Vocal Emotion Perception in non-professional Singers and Instrumentalists

Masterarbeit zur Erlangung des akademischen Grades
„Master of Science (M.Sc.)“

eingereicht bei

Dr. Christine Nussbaum
Professur für Allgemeine Psychologie I
am Institut für Psychologie

an der

Fakultät für Sozial- und Verhaltenswissenschaften
der Friedrich-Schiller-Universität Jena

Erstbetreuung:	Dr. Christine Nussbaum
Zweitbetreuung:	Prof. Dr. Stefan R. Schweinberger
Abgabedatum:	15. Juli 2024

Eingereicht von: Jessica Senftleben
Augustinerstraße 8, 99084 Erfurt
jessica.senftleben@gmail.com
Kognitive Psychologie und Kognitive Neurowissenschaften (M.Sc.)
153983

Contents

List of Figures

List of Tables

Abstract

In previous research, musicality has been shown to benefit vocal emotion perception. This is likely due to an increased auditory sensitivity to acoustic cues, such as fundamental frequency (F0) and timbre, among musicians. However, it is unclear how musicians with different forms of engagement or different levels of expertise differ from each other in their use of these acoustic cues to infer emotional meaning from speech prosody.

The present study aimed to investigate this by comparing the vocal emotion perception of singers (N = 45) and instrumentalists (N = 43) as well as of professional musicians (N = 40), non-professional musicians (amateurs, N = 88) and non-musicians (N = 38). For that purpose, we implemented an emotion recognition task with vocal stimuli that where manipulated to convey emotions either in all acoustic cues or selectively in F0 or timbre only.

Singers and instrumentalists did not significantly differ in their vocal emotion perception performance in any condition. The comparison of professional musicians, amateurs and non-musicians revealed that amateurs performed equally to professionals in full as well as F0- and timbre-modulated conditions. Furthermore, they performed better than non-musicians in a condition with full emotion modulation, but not in an F0-modulated condition. In both these conditions, professionals outperformed non-musicians. There were no differences between the groups in a timbre-modulated condition. Additionally, exploratory analyses revealed a link between dynamic pitch perception in music and speech that remained when controlling for music training.

These findings indicate that musicality benefits vocal emotion perception irrespective of different forms of musical engagement (singing vs. playing an instrument) or musical expertise (amateur vs. professional level). Musicality and enhanced vocal emotion perception appear to be linked by a naturally predisposed auditory sensitivity that particularly concerns pitch contours.

Zusammenfassung

Vorangegangene Untersuchungen haben gezeigt, dass sich Musikalität positiv auf die Wahrnehmung von Emotionen in der Stimme auswirkt. Dies ist wahrscheinlich auf eine erhöhte Sensibilität für akustische Parameter wie die Tonhöhe und Klangfarbe bei Musiker:innen zurückzuführen. Es ist jedoch unklar, wie sich verschiedene Musiker:innen in ihrer Nutzung dieser akustischen Parameter voneinander unterscheiden, um Emotionen in der menschlichen Stimme zu erkennen. Diese Nutzung könnte beispielsweise mit dem Level ihrer Expertise oder der Spezialisierung auf Gesang oder Instrumentenspiel variieren.

In der vorliegenden Studie wurden daher die Erkennungsleistungen einer Gruppe Sänger:innen (N = 45) mit einer Gruppe Instrumentalist:innen (N = 43) verglichen, sowie die Erkennungsleistungen von professionellen Musiker:innen (N = 40), Amateurmusiker:innen (N = 88) und Nicht-Musiker:innen (N = 38). Zu diesem Zweck implementierten wir eine Emotionserkennungsaufgabe, bei der Stimmaufnahmen so manipuliert wurden, dass Emotionen entweder in allen akustischen Parametern oder nur in der Tonhöhe oder Klangfarbe ausgedrückt wurden.

Sänger:innen und Instrumentalist:innen unterschieden sich in keiner der Bedingungen in ihrer Erkennungsleistung. Beim Vergleich der professionellen, Amateur- und Nicht-Musiker:innen zeigten Amateure vergleichbare Erkennungsleistung wie professionelle Musiker:innen, wenn die Emotionen durch alle Parameter und nur durch die Tonhöhe oder Klangfarbe ausgedrückt wurden. Außerdem zeigten Amateure bessere Erkennungsleistungen als Nicht-Musiker:innen, wenn die Emotionen durch alle Parameter ausgedrückt wurden, aber nicht, wenn sie nur durch die Tonhöhe ausgedrückt wurden. Professionelle Musiker:innen zeigten bessere Erkennungsleistungen als Nicht-Musiker:innen in beiden dieser Bedingungen. Die Gruppen unterschieden sich in ihren Erkennungsleistungen nicht, wenn nur die Klangfarbe emotional informativ war. Darüber hinaus ergaben explorative Analysen einen Zusammenhang zwischen der dynamischen Tonhöhenwahrnehmung in Musik und Sprache, der unabhängig von musikalischen Ausbildung bestehen blieb.

Diese Ergebnisse deuten darauf hin, dass Musikalität die Emotionserkennung in Stimmen begünstigt, und zwar unabhängig von verschiedenen Formen des musikalischen Engagements (Singen vs. Spielen eines Instruments) oder der musikalischen Expertise (Amateur- vs. professionelles Niveau). Musikalität und verbesserte Emotionserkennung in Stimmen scheinen durch eine Prädisposition zu auditiver Sensibilität verbunden zu sein, die sich insbesondere auf die Tonhöhe/Melodie bezieht.

1. Introduction

1.1 Vocal emotion perception

Emotions and their communication are an integral part of human behavior and social interaction. One way to express those emotions is the human voice. While emotions are of course conveyed through the verbal contents of human speech, they are also being communicated non-verbally. This happens via emotional speech prosody or via non-verbal vocalizations (e.g., laughter, moans, or cries; Pell et al., 2015). This work will focus on emotional prosody, which is the tone of voice of a speaker. As such it provides important information, as it matters greatly whether, for example, somebody calls for you in a happy or an angry tone (Brück et al., 2011). More precisely, speech prosody constitutes the modulation of suprasegmental properties of spoken language (Banse & Scherer, 1996; Grandjean et al., 2006). The properties (i.e., acoustic cues) likely to be most relevant for vocal emotion communication include perceived pitch, timbre, loudness, and tempo (Brück et al., 2011; Johnstone & Scherer, 2000; Juslin & Laukka, 2003).

The perceived pitch of a voice relates to the **fundamental frequency (F0)** of the vocal fold vibrations of a speaker. F0 varies continuously, depending on the length, tension, and mass of the vibrating body, i.e. the strings of a violin or an individual's vocal folds (Lagefoged, 1996). Due to a generally greater mass of vocal folds, the average F0 of male speakers is lower than of female speakers, with neutral voices ranging between 100 and 120 Hz in male and about 200 to 240 Hz in female speakers. For emotional voices, especially happiness, the F0 can be much higher (Nussbaum, Schirmer & Schweinberger, 2022). Comparatively, the modern international tuning standard for the musical note A_4 is a frequency of 440 Hz (International Organization for Standardization, 1975). The variation of F0 over time results in a dynamic pitch contour, which in both music and voice is called a melody. **Timbre** refers to the perceived quality of a voice or musical sound. It distinguishes between different instruments or voices in the absence of other prosodic differences, e.g. whether a specific melody is played on a violin or a clarinet. The same instrument or voice can also vary in its perceived timbre, depending on different playing or singing/speaking techniques (Gabrielsson & Juslin, 1996). Acoustic features like formant frequency, high-frequency energy, or the glottal waveform contribute to the perception of timbre and determine if a tone of voice is for example perceived as sharp or soft (Juslin & Laukka, 2003). The perceived loudness of a voice or sound corresponds to its **amplitude**. It is also often referred to as the intensity or energy of a voice (Banse & Scherer, 1996; Brück et al., 2011; Juslin & Laukka, 2003). The

temporal aspects of a voice or sound include the overall duration, speech rate (e.g., words per minute) and pauses. As a musical term, tempo describes how fast a piece is supposed to be played (e.g., beats per minute).

It has been hypothesized that acoustic cues follow distinct patterns to express different vocal emotions (Scherer, 1986, 2018). Indeed, there is some evidence to support this claim. Sadness has been described with a lower- to middle-ranged pitch, a soft timbre, a low amplitude and slow speech rate, whereas anger, joy and fear are often found to be characterized by a high pitch, high amplitude and fast speech rate (Banse & Scherer, 1996; Juslin & Laukka, 2003). Listeners generally exhibit above-chance performance when discriminating discrete emotions, even beyond the mentioned general patterns (Scherer, 1986, 2018). Research to further identify the distinct patterns differentiating between discrete emotions, however, has been inconclusive and heterogeneous. Results for fear and happiness were less consistent than those for anger and sadness, e.g. for pause distribution and proportion and formant frequencies (Juslin & Laukka, 2003).

While the discussed acoustic cues are all relevant to reliably perceive vocal emotions, they may contribute differently to the matter. Amplitude, speech rate, and especially pitch have been hypothesized to be the dominant cues (Banse & Scherer, 1996; Johnstone & Scherer, 2000). Timbre, while often mentioned and considered alongside these cues, has only recently been discussed in more detail (Anikin, 2020; Grichkovtsova et al., 2012; Nussbaum, von Eiff, et al., 2022; Piazza et al., 2018; Von Eiff et al., 2022). For instance, in a study by Nussbaum, Schirmer and Schweinberger (2022), timbre was found to be more important than pitch when recognizing pleasure, although this was not replicated in another study (Nussbaum, Schirmer & Schweinberger, 2023b).

Vocally expressed emotions are recognized across cultures (Laukka et al., 2016; Scheiner & Fischer, 2011; Scherer, 2018; Scherer et al., 2001) and most humans are able to efficiently and often automatically perceive them (Bigand et al., 2005; Chartrand et al., 2008; Lima et al., 2019; Thompson, 2009). This holds even in the absence of semantic content (e.g., Castro & Lima, 2010) and for languages unknown to the listener (Pell et al., 2009; Thompson & Balkwill, 2006). However, despite humans being considered experts for vocal emotions, there has been research indicating substantial interindividual differences in vocal emotion perception skills (Mill et al., 2009; Schirmer et al., 2005).

In line with recent scientific questions, this thesis will focus specifically on interindividual differences in vocal emotion perception skills related to the musicality of a person. For this purpose, the following section will first discuss the shared aspects of music and the human voice, followed by questions regarding interindividual differences due to nature vs. nurture. Then, the influence of musical expertise and different forms of musical engagement (singing vs. playing an instrument) will be addressed.

1.2 Musicality and vocal emotion perception

1.2.1 What is shared between music and voice?

Communication of emotions is not only significant in everyday spoken conversations, but also an important feature of art forms like music. Extensive scientific research has found and discussed a number of parallels, overlaps and interconnections in the emotional communication of music and voice.

Both speech and music are complex auditory stimuli that rely on acoustic cues to communicate emotions. The discussed acoustic cues of vocal stimuli that are involved when conveying emotion within speech (i.e., pitch, timbre, amplitude, tempo) apply similarly to the conveying of emotion in music (Coutinho & Dibben, 2013; pitch: Curtis & Bharucha, 2010; tempo, amplitude: Ilie & Thompson, 2006; overview: Juslin & Laukka, 2003; Scherer, 1995). In an extensive review, Juslin and Laukka (2003) concluded that there may be similar "emotion-specific patterns of acoustic cues" (p. 799) that are used when communicating emotions. In the studies they reviewed, high pitch, fast tempo and high intensity/amplitude were found to be associated with happiness, anger and fear, while sadness and tenderness were associated with low pitch and slow tempo, both in speech and music (Juslin & Laukka, 2003). Similarities were also found regarding recognition patterns. Within vocal stimuli, anger is often recognized better than other emotions (e.g., Banse & Scherer, 1996; Thompson & Balkwill, 2006), which pertains to music as well, albeit not to the same extent (Juslin & Laukka, 2003).

Pitch, temporal aspects and instrumentation are considered integral when composing music (Juslin & Laukka, 2003; Schutz, 2017), although the exact importance of these individual acoustic cues for musical emotion production is hard to pinpoint. The choice of different instruments with their unique acoustic restraints and possibilities as well as the myriad of music styles and genres result in a vast variety of ways to communicate emotions (Schutz, 2017). Additionally, musical emotion expression is not only dependent on decisions of the composer about harmony or instrumentation but also influenced by the performer and how they perform a music piece and its emotions. This is a marked difference to vocal emotion expression, where the differentiation of composer vs. performer usually does not exist. Another difference lies in the unique capabilities of voices and music. Instruments can be limited in their way of communicating emotional meaning due to their individual acoustic restraints. For example, a banjo is difficult to play in a slow tempo and/or low pitch (Stephey & Moore, 2008) and as such has an impaired ability to convey sadness (Schutz, 2017). The human voice, in turn, has its own restraints and is not able to achieve certain acoustic features that are important in music, such as modality and harmonic progression (Juslin & Laukka, 2003). Specifically the modality (minor vs. major chords) is strongly associated with emotional meaning in Western music (Schutz, 2017), with major modes often perceived as happy and minor modes as sad or fearful (Gagnon & Peretz, 2003; Quinto & Thompson, 2012).

Another overlap of music and voice is reflected in the neural processing of emotional music and voices, as both probably recruit similar neural circuits (Aubé et al., 2015; Escoffier et al., 2013; Frühholz et al., 2016). Frühholz et al. (2016) propose a neural network which includes the amygdala, the auditory cortex, the insula, the cerebellum and the inferior frontal cortex. Indeed, neuroimaging studies found shared neural codes in the auditory cortex for the emotional processing of instrumental music cues and vocal cues (Paquette et al., 2018; Sachs et al., 2018). Neural activity in the amygdala in response to fearful (instrumental) music and vocalizations were found to be correlated (Aubé et al., 2015).

Vocally expressed emotions are recognized across cultures (Laukka et al., 2016; Scheiner & Fischer, 2011; Scherer, 2018; Scherer et al., 2001), which is also the case for musical emotions (Fritz et al., 2009; Thompson & Balkwill, 2010). While there seem to be culturally nuanced schemata for both recognition and expression of emotions in music and voice (Laukka et al., 2016; Thompson & Balkwill, 2010), emotion recognition appears to be innate to a certain degree (Scheiner & Fischer, 2011).

There are several possible interpretations for the similarities between voice and music regarding their emotion communication. For it to be merely coincidental seems unlikely given the large number of similarities in acoustic cues (Juslin & Laukka, 2003), the overlap in their neural circuits (Frühholz et al., 2016) and the cross-cultural evidence. One theory takes the view that the similarities are a result of music and voice evolving in parallel from a common origin (Darwin, 1871, 1872). The common origin is presumed to be a pre-linguistic system of communication, often termed “protolanguage” (Bickerton, 1990; Fitch, 2006, 2013), which was used in territoriality and courtship, for emotion expression (Darwin, 1871; Fitch, 2013;

Thompson et al., 2012) and to maintain parent-infant bonds (Dissanayake, 2000). Another theory proposes the human voice as a model for music (Brown, 2000; Juslin & Laukka, 2003; Mithen et al., 2006). Here, musical emotion expression evolved as an imitation of (non-verbal) vocal emotion expression, possibly concurrently with language (Brown, 2000; Mithen et al., 2006).

Regardless of the exact evolutionary origins, it can be concluded that a lot is shared between music and voice in respect of their emotional communication, with findings that suggest parallels in acoustic patterns, overlaps of neural circuits and cross-cultural emotion recognition.

1.2.2 Role of individual differences

While musical and vocal emotions are recognized consistently and quickly across listeners (Bigand et al., 2005; Scherer et al., 2017; Thompson, 2009), there can be substantial interindividual differences. One reason for such differences may be the musicality of listeners. Standing to reason, musicality was found to be positively associated with an enhanced processing of musical emotions (Bhatara et al., 2011; Castro & Lima, 2014; Nolden et al., 2017; Sharp et al., 2019). For vocal emotions, research has found similar evidence (M. Martins et al., 2021; Nussbaum & Schweinberger, 2021). How exactly and why musicality is related to an enhanced vocal emotion perception is a matter of debate.

It has been argued that musicality is associated with an enhanced sensitivity to acoustic cues (Besson et al., 2011; Kraus & Chandrasekaran, 2010; Shorey et al., 2024), which due to the discussed similarities in acoustic patterns of vocal and musical emotions results in better vocal emotion perception (Nussbaum & Schweinberger, 2021). In contrast to this, a study by Trimmer and Cuddy (2008) found that participants' musical training and musical perception abilities did not predict emotion perception performance. Instead there was an association with general emotional intelligence, which suggests that enhanced emotion recognition in music and voice may be a result of the operations of a "high-level, perceptual emotional processor" rather than similar processes at a perceptual level (Trimmer & Cuddy, 2008, p. 847). While this theory has some merit, it stands in stark opposition to other findings, notably those of Correia et al. (2022), where auditory perception skills fully mediated the link between musicality and vocal emotion perception. Further, neuroimaging research comparing musicians and non-musicians suggests that there are differences in brain responses in the early stages of vocal emotion processing (Pinheiro et al., 2015; Rigoulot et al., 2015), which relate to basic acoustic analysis (Schirmer & Kotz, 2006). Therefore, it seems plausible for auditory sensitivity to be relevant for the link between musicality and vocal emotion perception.

It is unclear to which degree the link between musicality and enhanced emotion recognition is related to environmental factors, such as music training, and what role genetic factors, such as talent, play (i.e., nature vs. nurture). Regarding the argument for **natural talent**, a person could just have an innate capacity for perceiving both musical and vocal acoustic cues. According to the protolanguage hypothesis, music and speech once evolved from the same origin (Darwin, 1871, 1872; Fitch, 2013) and as such musical and vocal emotion processing capacities could be linked through genetic factors. Evidence from cross-cultural studies also supports the notion of an innate capability for the perception of vocal and musical emotions (Scherer et al., 2001; Thompson & Balkwill, 2010). A review by Trehub (2003) discusses infant-adult parallels in music processing and comes to the conclusion that the foundations of music perception lie in genetic rather than cultural factors. Arguments supporting the **nurture side** state that interindividual differences are caused by training. Because the acoustic cues of musical and vocal emotions are so similar, musical training and repeated engagement with musical cues may also improve capabilities in vocal emotion perception. The OPERA hypothesis (Patel, 2011, 2014) postulates that benefits of music training only transfer to other domains when certain conditions are met: There needs to be an *overlap* of shared neural networks involved in sound processing in music and speech, and music has to demand more processing *precision* from the shared networks than speech. Further, musical activities have to induce positive *emotion*, be frequently *repeated* and be paid focused *attention* (Patel, 2011). When music training meets these conditions, it results in an adaptive neural plasticity that leads to higher processing precision, which is needed for music and through the shared neural networks also benefits speech processing. Sufficient repetition likely takes a long time: it often takes years of music training to elicit observable changes in plasticity (Kraus & White-Schwoch, 2017). Neural sound encoding has also been found to be positively correlated with the number of years of music training (see Kraus & Chandrasekaran, 2010). This supports the notion that training is at least partially responsible for interindividual differences in vocal sound processing. Alternatively, the link between enhanced emotion recognition and musicality could be the result of a **combination of nature and nurture**. Similar to musicality, it may be that a predisposition (i.e., innate auditory sensitivity) *and* the opportunity to establish these capabilities (i.e., music training) are necessary. For musical abilities such as absolute pitch, this has been argued (Baharloo et al., 1998; Zatorre, 2003). How much genetics and environmental factors each contribute to the process remains unclear due to their interactions.

An interesting point in this debate are individuals with congenital amusia. Congenital amusia is a disorder where individuals suffer no hearing loss or other pertaining cognitive deficits while having severe problems with music perception and production since birth, particularly with pitch processing (Peretz, 2008; Peretz & Hyde, 2003). Studies found that the majority of individuals with amusia have no problems with perceiving pitch in speech (Ayotte et al., 2002; Patel et al., 2008), which could indicate a dissociation of musicality and the processing of speech prosody. Concerning emotion perception, however, participants with amusia performed poorer than controls when recognizing emotions in speech (Cheung et al., 2021; Lima et al., 2016; Thompson et al., 2012; Y. Zhang et al., 2018), which may be associated with a particular deficit in pitch perception (Lolli et al., 2015; Pralus et al., 2019). Taken together, these findings indicate that deficits in vocal emotion perception are linked to congenital deficits in music perception and as such suggest mediating genetic factors for the link between vocal emotion perception and musicality (Nussbaum & Schweinberger, 2021).

1.2.3 Musicians vs. non-musicians

One method to study the effects of musicality and musical activities and how they extend to non-musical domains and abilities is the comparison of musicians and non-musicians. In that regard, musicians are often defined as having received a certain amount of music training or music lessons in their life, usually at least several years (J. D. Zhang et al., 2020). Current, regular musical activity, such as practice, or a training onset in childhood are other common criteria (Nussbaum & Schweinberger, 2021). Non-musicians as controls are required to not engage in musical activities like this, even though they may engage in different ways (e.g., listening to music).

As mentioned, musicality has been linked to benefits in auditory processing, so far as to consider musicians as particularly sensitive to acoustic cues. Regarding speech prosody, this **auditory expertise** concerns especially vocal pitch (Magne et al., 2006; Marques et al., 2007; Schön et al., 2004), but also timbre (Chartrand & Belin, 2006; Münzer et al., 2002) and timing (Sares et al., 2018). In addition to that, enhancements were found for a range of perception and language abilities, for example recognizing speech-in-noise, segmental processing of speech sounds, and tracking of language (Coffey et al., 2017; Elmer et al., 2018; Hallam, 2017; Schellenberg & Lima, 2024; Strait & Kraus, 2011). Moreover, musicians’ benefits have been observed for general **cognitive abilities**, such as long-term, short-term and working memory, executive functions, intelligence and academic achievement (Schellenberg, 2016; Schellenberg & Lima, 2024; Talamini et al., 2017). Beyond cognitive functioning, music training is also a popular model for studying **brain plasticity** (Habib & Besson, 2009; Herholz & Zatorre, 2012; Moreno & Bidelman, 2014; in opposition: Schellenberg & Lima, 2024), due to observed neuroanatomical and -physiological changes linked to musical expertise.

Regarding **vocal emotion perception** evidence is limited and quite heterogenous (M. Martins et al., 2021; Nussbaum & Schweinberger, 2021). One of the first studies by Nilsonne and Sundberg (1985) found music students to be better than law students when determining whether voice recordings were made during a depressive phase of the speaker or not. Similar results were found by Thompson et al. (2004), where musically trained adults performed better than untrained adults when identifying sadness, fear and neutral emotions. Confounding these results, however, music training was correlated with general cognitive abilities. Nevertheless, other studies found musicians to be more accurate than non-musicians even when controlling for general cognitive abilities, personality traits, and socio-educational background (Correia et al., 2022; Lima & Castro, 2011; Nussbaum, Schirmer & Schweinberger, 2023b). In contrast to this evidence, a number of studies reported null findings (Başkent et al., 2018; Mualem & Lavidor, 2015; Trimmer & Cuddy, 2008).

The previously discussed notion of auditory expertise and sensitivity leaves questions about the exact acoustic processing differences between musicians and non-musicians. It could be that musicians are better at perceiving specific acoustic cues, leading to a cue-specific performance improvement. Research in that line particularly proposes pitch contour (F0) as important in vocal emotion perception (e.g., Fuller et al., 2014; Globerson et al., 2013; Nussbaum, Schirmer & Schweinberger, 2023b). Alternatively to cue-specific processing, musicians may be able to process acoustic cues in general more efficiently, e.g. faster or by more selectively attending to relevant cues (Symons & Tierney, 2024), which improves vocal emotion perception (Lima & Castro, 2011; Nussbaum & Schweinberger, 2021).

Another question arises whether listening passively as opposed to engaging actively with music makes a noticeable difference for vocal emotion perception. While passively listening to music may be sufficient to develop a certain level of musicality (Bigand & Poulin-Charronnat, 2006), several studies suggest that an active engagement with music (e.g., multimodal sensorimotor-auditory music training) is essential to acquire the fine-tuned auditory perception skills needed for vocal emotion perception (Chari et al., 2020; Fuller et al., 2018). The same pertains to changes in neuroplasticity associated with musicality (Kraus & White-Schwoch, 2017; Lappe et al., 2011; Palomar-García et al., 2017; Pantev & Herholz, 2011). Furthermore, in a recent study by Toh et al. (2023), active musicians were better than former musicians in a prosodic pitch perception task, and former musicians in turn outperformed non-musicians. This suggests active engagement to be advantageous over former activity, even with a comparable amount of (multimodal) music training.

The overall heterogeneity in findings could be due to methodological limitations, such as confounding variables (Lima & Castro, 2011; Thompson et al., 2004) or small sample sizes (e.g., Başkent et al., 2018; Nolden et al., 2017; Park et al., 2015; Pinheiro et al., 2015; Weijkamp & Sadakata, 2017). If the effect of musicality is a small one (e.g., Correia et al., 2022; Fuller et al., 2014), it would need large samples to emerge (M. Martins et al., 2021). It is further important to keep in mind that most studies comparing musicians and non-musicians discussed here are correlational in design (Schellenberg, 2020a), which makes reliable conclusions about causality impossible. Whether an enhanced vocal emotion perception is the result of musicality and/or musicality is the results of an enhanced vocal emotion perception is not unequivocally ascertainable. Correlational and quasi-experimental designs further include the possibility of pre-existing interindividual differences regarding auditory sensitivity, personality, cognitive abilities or socioeconomic status, which could inform whether and how long participants receive music training (see e.g., Corrigall et al., 2013) and how they perform in non-musical domains. In a recent review, Schellenberg and Lima (2024) argue that while correlational data suggest positive associations, longitudinal studies provide little to no evidence for a causal link from music training to enhanced listening skills or to enhancements in non-musical domains such as the perception of emotional prosody or speech-in-noise, language skills, memory, executive functions, and general cognitive ability. Indeed, longitudinal evidence about the effectiveness of music training interventions on vocal emotion perception is sparse and not without methodological problems. For example, Thompson et al. (2004) tested six-year-old children after randomly assigning them to one year of either keyboard, singing, drama, or no lessons. While the keyboard and drama group performed better than the no-lesson group, the singing group did not. Noteworthily, less than 30 % of the initially recruited children returned for testing (Schellenberg & Lima, 2024), which leaves questions about a possible sample bias. Other studies found musical interventions to improve emotion recognition in adults (Bodner et al., 2012; Mualem & Lavidor, 2015), but as the interventions explicitly focused on emotion identification, their effectiveness could be due to "training to the test" (Nussbaum & Schweinberger, 2021).

Schellenberg and Lima (2024) conclude that rather than music training and transfer effects, pre-existing factors, e.g. perceptual skills, may account for positive associations between musicality and non-musical abilities. While this likely pertains to non-auditory cognitive functioning (Sala & Gobet, 2020; Schellenberg, 2016), training effects cannot be completely ruled out for all non-musical domains. Differences in brain responses, speech processing and auditory sensitivity between musicians and non-musicians suggest at least some effects of music training (Elmer et al., 2018; Kraus et al., 2014; Kraus & Chandrasekaran, 2010; Neves et al., 2022).

As mentioned above, research comparing musicians and non-musicians often defines musicians as having received some amount of music training (J. D. Zhang et al., 2020). The distinction between professional musicians, whose career involves music (e.g., as members of professional orchestras or choirs), and non-professional musicians (i.e., amateurs), who are musically trained but do not work as musicians, is rarely made (Vincenzi et al., 2022). In the existing literature, amateur musicians scored higher when tested for general cognitive ability (Vincenzi et al., 2022) and had a lower “brain age” (Rogenmoser et al., 2018) in comparison to professionals. Furthermore, neuroimaging studies found increased activation in professional singers and instrumentalists when compared with corresponding amateurs during singing or imagined musical performance (Kleber et al., 2010; Lotze et al., 2003), which may reflect experience-dependent changes in neural audio-motor processes. Taken together, these findings indicate qualitative differences between professional musicians and amateurs. However, it is unclear whether the groups differ in regard to vocal emotion perception. In light of evidence that predisposed auditory sensitivity rather than the amount of music training influences the link between musicality and vocal emotion perception (Correia et al., 2022; Nussbaum, Schirmer & Schweinberger, 2023b), the possibly more extensive training and/or practice regime of professional musicians may not result in an advantage over amateur musicians. Further research is needed to gather conclusive evidence.

1.2.4 Singers vs. instrumentalists

Another distinction rarely made when considering musicians as a group in research on vocal emotion perception is the one between singers and instrumentalists. Singing vs. playing an instrument require distinct sets of motoric expertise and auditory-motor skills (Fisher et al., 2020; Krishnan et al., 2018), with singers predominantly making use of their vocal tract, while instrumentalists especially employ fine motor skills in their hands. Both skill sets are required for brass and woodwind players. Neuroimaging studies documented sensorimotor brain responses that related to specific musical expertise, e.g. larger cortical representation of the fingers on the left hand in string players than in non-musicians (Elbert et al., 1995) or expertise-selective enhancements for beatboxing vs. playing guitar (Krishnan et al., 2018). Halwani et al. (2011) found structural differences between singers, instrumentalists and non-musicians, with singers showing increased volume and microstructural complexity in the arcuate fasciculus, a white matter tract connecting regions involved in sound perception and production. Related, there is evidence for advantages of singing over playing an instrument in vocal production. Singers outperformed instrumentalists in imitation of foreign speech (Christiner & Reiterer, 2015), which implies adaptive plasticity (for speech imitation) to not rely on auditory processes alone but also on vocal-motor induced processes.

Regarding the processing of specific acoustic cues, findings suggest singers and instrumentalists to be similar in their advantage over non-musicians in pitch perception. In studies by Nikjeh et al. (2008, 2009), singers and instrumentalists outperformed non-musicians in pitch discrimination and pitch production accuracy, while there were no significant performance differences between the two musician groups. Electroencephalographic (EEG) responses indicated that singers with instrumental training may have an auditory neural advantage over vocal- or instrumental-only musicians (Nikjeh et al., 2008), but this was not reflected in the behavioral measures (Nikjeh et al., 2009). Similarly, singers and instrumentalists outperformed non-musicians, but did not perform significantly differently from each other when perceiving stimuli in non-native tonal languages (Kirkham et al., 2011; Rø et al., 2006). Comparing instrumentalists, studies by Rauscher and Hinton (2011) found better pitch discrimination in string players, whereas percussionists performed better in temporal discrimination, which indicates processing advantages for acoustic cues respective to instrument-typical demands. Specifically concerning timbre, there seem to be processing differences between different instrumentalists. Violinists, trumpeters and flutists showed enhanced auditory cortical representation for musical timbres that related to their principal instrument when compared to the timbre of other instruments (Margulis et al., 2009; Pantev et al., 2001). Similarly, pianists and violinists exhibited enhancements specifically to tone stimuli of their respective instruments in an EEG study by Shahin et al. (2008). This raises the possibility of singers being tuned particularly to vocal timbre. However, these processing advantages for familiar timbres may not extend to behavioral measures. Comparing flutists, trumpeters and violinists, Holmes et al. (2022) found no instrument-specific improvement in pitch discrimination performance. Weiss, Vanzella and colleagues (2015) expected pianists to have an advantage when remembering piano melodies in comparison to melodies on a banjo or marimba. The results showed no such advantage, which indicates that processing advantages do not extend to memory effects.

Empirical evidence on differences between singers and instrumentalists for vocal emotion perception is limited and inconclusive. On the one hand, singing is a form of musical expression that is arguably the closest to vocal emotion expression, which could make singing more advantageous over instrumental musical activities for vocal emotion perception. In line with this, Greenspon and Montanaro (2023) found that the objectively measured singing ability predicted vocal emotion recognition performance in a sample of college students, while pitch discrimination ability or self-reported musical experience did not. Corroborating this finding, exploratory analyses found positive correlations of self-rated singing abilities with vocal emotion perception performance (Correia et al., 2022; Nussbaum, Schirmer & Schweinberger, 2023b). On the other hand, singing lessons could interfere with vocal emotion perception, for example due to possible conflicts of cue application in singing vs. natural prosodic use in normal speech. In that line stands the intervention study by Thompson et al. (2004) where children performed better in vocal emotion perception after a year of keyboard lessons, which was not found for singing lessons. The authors consider that keyboard lessons may produce more precise pitches and pitch intervals than singing lessons with children typically involve. A third possibility lies in a general musicality benefit to vocal emotion perception, with no discernible differences between singers and instrumentalists. Supporting that notion, a recent EEG study found no differences in the processing of vocal or musical emotions related to singing vs. playing an instrument (I. Martins et al., 2022). Notably, a master thesis by Francisco (2021) found no significant differences in vocal emotion perception performance between singers and instrumentalists. However, there were also no differences when comparing the two groups to non-musicians, and the sample size was small (9 to 13 participants per group). In light of the assumed importance of pitch for vocal emotion perception (Banse & Scherer, 1996; Johnstone & Scherer, 2000) and similar performances in pitch discrimination tasks by singers and instrumentalists (Nikjeh et al., 2008, 2009), these findings may suggest musicality to be beneficial to vocal emotion perception irrespective of a specific expertise for an instrument or singing.

1.3 Present study and hypotheses

Previous work has shown that musicians outperform non-musicians in vocal emotion recognition, which was recently replicated in a study comparing professional musicians with non-musicians (Nussbaum, Schirmer & Schweinberger, 2023b). In that study it was further shown that musicians are particularly sensitive to emotional pitch cues in voices, which seems to be due to a predisposed natural sensitivity towards pitch and melody cues rather than musical training.

The present study was designed as a follow-up on Nussbaum, Schirmer and Schweinberger (2023b) and addresses two main objectives: the first was a comparison of the vocal emotion perception of two groups of non-professional musicians, by recruiting a well-powered sample with singers and instrumentalists. Currently, the available literature on singers and instrumentalists with regard to vocal emotional processing is inconsistent and requires further systematic research. Second, the present study assessed specifically non-professional musicians (amateurs), in contrast to the professional ones recruited by Nussbaum, Schirmer and Schweinberger (2023b). There is accumulating evidence for qualitative differences between professional and non-professional musicians. However, it is unclear how these groups diverge with regard to vocal emotion perception. Therefore, we pooled the data collected in the present study and in Nussbaum, Schirmer and Schweinberger (2023b) to compare the performance of musical amateurs to the one of professional musicians and non-musicians.

In their study, Nussbaum, Schirmer and Schweinberger (2023b) employed parameter-specific voice morphing to investigate the particular role of pitch and timbre for vocal emotion perception. Voice morphing as a tool allows for the manipulation of vocal stimuli in such a way that emotions are expressed only through specific acoustic cues, with the other cues rendered emotionally uninformative (Kawahara et al., 2008; Kawahara & Skuk, 2018). Thus, in line with Nussbaum, Schirmer and Schweinberger (2023b), we implemented the same manipulated vocal stimuli expressing emotions either through pitch or timbre only.

In consideration of existing evidence, we expected no differences between singers and instrumentalists in their overall vocal emotion recognition performance or in their vocal emotion recognition performance based on pitch and timbre cues only. Further, we expected non-professional musicians to perform better than non-musicians and equally to or better than professional musicians in their overall vocal emotion recognition performance and in a condition when only pitch was emotionally informative. We expected no differences between non-professional musicians, professional musicians and non-musicians in vocal emotion recognition performance based on timbre cues only.

2. Methods

The present study was designed as a follow-up and extension of the study by Nussbaum, Schirmer, and Schweinberger (2023b). The design is therefore nearly identical to their study, except that a different sample was used. The data of professional musicians and non-musicians was re-used from Nussbaum, Schirmer and Schweinberger (2023b). The sample of non-professional musicians (singers vs. instrumentalists) was recruited additionally for the purpose of the present study. Note that in Nussbaum, Schirmer and Schweinberger (2023b), all participants were invited for a subsequent EEG session after completing the online experiment, which is reported in Nussbaum, Schirmer and Schweinberger (2023a). For the additional amateur sample, this was not the case. Any further methodological differences and changes to Nussbaum, Schirmer and Schweinberger (2023b) are explicitly stated in this methods section.

2.1 Participants

The sample size of 39 professional musicians and 38 non-musicians in the study by Nussbaum, Schirmer and Schweinberger (2023b) was sufficiently powered to reveal medium-sized group effects in the Full (d = 0.81, 95%-CI [0.31, 1.29]) and the F0 condition (d = 0.56, 95%-CI [0.07, 1.04]). Further, is was sufficiently powered to detect correlations r > .25. Therefore in the present study, we aimed at a sample size of 40 participants per group of non-professional musicians (singers vs. instrumentalists).

Data collection of professional musicians and non-musicians took place from June 2021 to May 2022 and is described in detail in Nussbaum, Schirmer and Schweinberger (2023b). Data of non-professional musicians were collected from June 2023 to January 2024. The data collection was pseudonymized. All participants were aged between 18 and 54 years and fluent German speakers. Participants provided informed consent before completing the experiment and received compensation in form of 12.50 € or course credit upon completion.

The experiments were in line with the ethical guidelines of the German Society of Psychology (DGPs) and approved by the local ethics committee of the Friedrich Schiller University Jena (Reg.-Nr. FSV 19/045).

Professional musicians Recruitment criteria specified professional musicians as having either a music-related academic degree or a non-academic music qualification. Data from 41 professional and semi-professional musicians (further also simply referred to as *professionals*) were collected by Nussbaum, Schirmer and Schweinberger (2023b). The data from two professionals had to be excluded because they omitted > 5 % trials in the emotion classification task. One participant's data collected for the group of non-professional musicians were included in the group of professionals, as the participant had a master's degree in music science. Thus, data from 40 professionals entered analysis (20 male, 20 female, aged 20 to 42 years [M = 29.6; SD = 5.58]). Mean onset age of musical training was 7 years (SD = 2.50, 4 – 17 years). Twenty-five participants were professional musicians with a music-related academic degree, all others had a non-academic music qualification (i.e., they worked as musicians or won a music competition; for more details see Table A.2). Thirty-six participants had studied their instrument for over 10 years, three between 6 - 9 years and one between 4 - 5 years.

Non-musicians Recruitment criteria specified that non-musicians had not learned an instrument and did not engage in any musical activities like choir singing during childhood. Nussbaum, Schirmer and Schweinberger (2023b) recorded data from 40 non-musicians, of which two exceeded the > 5 % omission criterion. Thus, data from 38 non-musicians were available for re-analysis (18 male, 20 female, aged 19 to 48 years [M = 30.5; SD = 6.54]; see Table A.2 for more demographic details).

Non-professional musicians I recorded data from 94 non-professional musicians (further also referred to as *amateurs*) that were divided into two groups. The groups will be referred to as *singers* and *instrumentalists* respectively.

Singers Recruitment criteria specified that singers had to be non-professional musicians (i.e., they held no music-related academic degree or worked professionally as a musician) who are currently active in a choir or another singing group. Further, they should not play an instrument actively and regularly (i.e., they must not currently be in an orchestra or a band). I recorded data from 48 singers, of which two were excluded because they omitted > 5 % trials in the emotion classification task. Another participant was excluded because the stimuli in the emotion classification task did not play correctly. Thus, data from 45 singers were analyzed (22 female, 22 male, aged 18 to 53 years [M = 27.02, SD = 8.2]). Mean onset age of musical training was 8 years (SD = 3.08, 5 - 20 years). Twenty-two participants had over 10 years of training, ten between 6-9 years, four between 4-5 years and three between 2-3 years. Six participants had less than one year of training (for more details see Table A.1).

Instrumentalists Recruitment criteria specified that instrumentalists had to be non-professional musicians (i.e., they held no music-related academic degree or worked professionally as a musician) who are currently active in an orchestra or a band. Further, they should not engage in singing activities actively and regularly (i.e., they must not currently be in a choir or another singing group). Data from 46 instrumentalists were collected, of which one instrumentalist was excluded because she was also currently active in a choir and another instrumentalist was shifted into the group of professionals because he had a Master's degree in music science. Another participant was excluded, because the stimuli in the emotion classification task did not play correctly. Thus, data from 43 instrumentalists entered analysis (24 female, 18 male, aged 18 to 54 years [*M* = 28.51, *SD* = 10.64]). Mean onset of musical training was 7 years (*SD* = 2.27, 4 - 14 years). Thirty-five participants had studied their instrument for over 10 years, one between 6-9 years and three between 1-2 years. Four participants had less than one year of training (for more details see Table A.1).

2.2 Stimuli

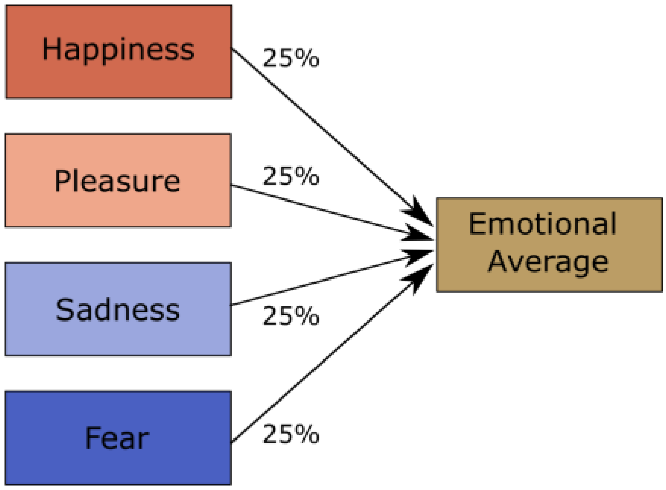
For the present study, stimuli are identical to the ones used in Nussbaum, Schirmer and Schweinberger (2023b; see https://doi.org/10.17605/OSF.IO/3JKCQ). Thus, the description of the stimulus (whole section 2.2) material is directly quoted from this study.

Original audio recordings We selected original audio recordings from a database of vocal actor portrayals provided by Sascha Frühholz, similar to the ones used in Frühholz et al. (2015). For our studies, we used three pseudowords (/molen/, /loman/, /belam/) uttered by eight speakers (four male, four female) with expressions of happiness, pleasure, fear, and sadness.

Voice averaging Using the Tandem-STRAIGHT software (Kawahara et al., 2008, 2013), Nussbaum, Schirmer and Schweinberger (2023b) created emotional averages from the four emotions used in the study (see Figure 1) for each speaker and pseudoword. These averages, although not neutral, were uninformative and unbiased with respect to the four emotions of interest. We opted for average rather than neutral stimuli because a previous study showed that averages are more suitable for the subsequent generation of voice morphs ensuring that such morphs do not differ systematically in perceived naturalness (Nussbaum, Pöhlmann, et al., 2023).

Figure 1

Schematic depiction of the voice averaging process



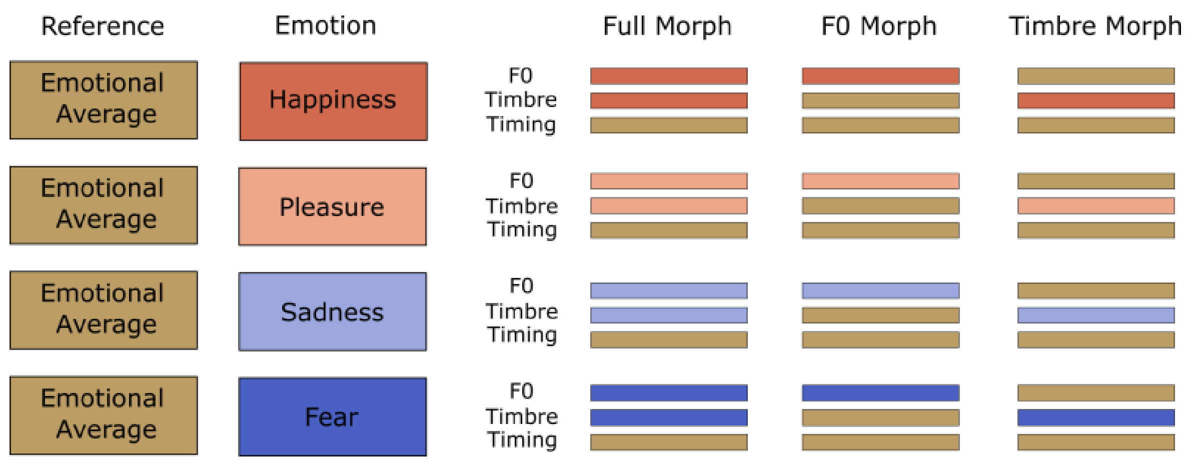
Note. This figure was taken from Nussbaum, Schirmer & Schweinberger (2023b), p. 6.

Parameter-specific voice morphing To synthesize parameter-specific emotional voice morphs, morphing trajectories between each emotion and the emotional average of the same speaker and pseudoword were created. After manual mapping of time- and frequency anchors at key features of a given utterance pair (e.g., on- and offset of vowels), vocal samples on an emotion/average-continuum were synthesized via weighted interpolation of the originals; for a more detailed description see Kawahara and Skuk (2018). Crucially, Tandem-STRAIGHT allows independent interpolation of five different parameters: (1) F0-contour, (2) timing, (3) spectrum-level, (4) aperiodicity, and (5) spectral frequency; the latter three are summarized as timbre.

Three types of morphed stimuli were created (see Figure 2). Full-Morphs were stimuli with all Tandem-STRAIGHT parameters taken from the emotional version (corresponding to 100 % from the emotion and 0 % from average), with the exception of the timing parameter, which was taken from the average (corresponding to 0 % emotion and 100 % average). F0-Morphs were stimuli with the F0-contour taken from the emotional version, but timbre and timing taken from the average. Timbre-Morphs were stimuli with all timbre parameters taken from the emotional version, but F0 and timing from the average. In addition, all average stimuli were included as a further ambiguous reference category. Note that the timing was kept constant across all conditions to allow a pure comparison of F0 vs. timbre. In total, this resulted in 8 (speakers) × 3 (pseudowords) × 4 (emotions) × 3 (morphing conditions) + 24 average (8 speakers × 3 pseudowords) = 312 stimuli. For analysis purposes, we collapsed data across speakers and pseudowords. Using PRAAT (Boersma, 2018), we normalized all stimuli to a root-mean-square of 70 dB SPL (duration *M* = 780 ms, range 620 to 967 ms, *SD* = 98 ms). Please refer to https://doi.org/10.17605/OSF.IO/3JKCQ for a detailed summary of acoustic parameters and some examples of the sound files.

Figure 2

Morphing matrix for stimuli with averaged voices as reference



Note. This figure was taken from Nussbaum, Schirmer & Schweinberger (2023b), p. 6.

2.3 Design

As previously stated, the design of the present study is very similar to Nussbaum, Schirmer and Schweinberger (2023b). For professionals and non-musicians, the study consisted of two sessions. They first completed an online session outside the laboratory, which was then followed by an EEG session in the laboratory. For amateurs, the study consisted of the online session only, followed by a short personal meeting for a debriefing. The meeting after the online session was included to make the study more comparable to the study by Nussbaum, Schirmer and Schweinberger (2023b), as the EEG session after the online session might have increased the commitment and conscientiousness in the online session. The post-experiment meeting in the present study was intended to have a similar effect. All participants were asked to ensure a quiet environment during the online session and to use a computer with a physical keyboard and headphones. For technical reasons, Google Chrome was recommended as browser for the online session while Safari and Firefox were excluded.

The online data collection was implemented via PsyToolkit (Stoet, 2010, 2017). At first, all participants were asked for demographic information (age, gender, native language, profession, and potential hearing impairments such as tinnitus). Amateurs were asked an additional question regarding instrumental and singing activities, i.e. if their musical activity consisted predominantly either of singing or playing an instrument. Then, participants could adjust their sound settings to a comfortable sound pressure level which they were asked not to change for the remaining duration of the study. Afterwards, participants performed the vocal emotion classification experiment and the PROMS test, which were both fully identical to Nussbaum, Schirmer and Schweinberger (2023b). Finally, they completed several questionnaires. Mean duration of the whole online experiment was about 75 minutes.

Vocal emotion classification experiment The experimental design was identical to Nussbaum, Schirmer and Schweinberger (2023b). Therefore, its description is directly quoted: The participants' task was to classify vocal emotions as happiness, pleasure, fear, or sadness. Each trial started with a green fixation cross presented for 500 ms. Subsequently, a loudspeaker symbol appeared, and the sound was played. After voice offset, a response screen showed the emotion labels and participants could enter their response within a 5000 ms time window starting from voice offset. Participants responded with their left and right index and middle fingers. The mapping of response keys to emotion categories was randomly assigned for each participant, out of four possible key mappings. Emotions of the same valence were always assigned to the same hand and emotions with similar intensity (fear – happiness and sadness – pleasure) were always assigned to the corresponding fingers of both hands (details in Tables A.5, A.6 and A.7). In case of no response (omission error), the final trial slide (500 ms) provided a feedback prompting participants to respond faster; otherwise, the screen turned black. Then the next trial started.

At the beginning of the experiment, participants completed eight practice trials with stimuli not used during the actual task. Subsequently, all 312 experimental stimuli were presented once in randomized order across six blocks of 52 trials each. Between blocks, participants could take self-paced breaks. The total duration of the experiment was about 25 minutes.

Profile of Music Perception Skills (PROMS) To measure music perception skills beyond self-reports, we adopted the modular version of the *Profile of Music Perception Skills* (PROMS; Law & Zentner, 2012; Zentner & Strauss, 2017). We selected the four subtests “Melody”, “Pitch”, “Timbre”, and “Rhythm”, which we considered most informative for the present research. For each subtest, participants completed 18 items, preceded by one practice trial. During each trial, participants heard a reference stimulus twice followed by a target stimulus. Then, they indicated whether reference and target were the same or different. Although this was a binary decision, the test employs a 5-point Likert scale with the labels “definitely same”, “maybe same”, “don’t know”, “maybe different”, and “definitely different”, which we also adopted here. Participants completed the test in about 20 minutes. One participant encountered technical problems in the “Melody” subtest, which was therefore repeated several months later to be included in data analysis.

Questionnaires After the PROMS, participants completed several questionnaires: the German Version of the *Autism Quotient Questionnaire* (AQ; Baron-Cohen et al., 2001; Freitag et al., 2007), a 30-item Personality Inventory measuring the Big Five domains (short form of the *Big Five Inventory-2*, BFI-2-S; Rammstedt et al., 2020), the *Goldsmiths Musical Sophistication Index* (Gold-MSI; Müllensiefen et al., 2014) to assess the participants’ degree of self-reported musical skills, additional questions concerning music experience and musical engagement, their socioeconomic background, and the 20-item version of the *Positive-Affect-Negative-Affect-Scale* (PANAS; Breyer & Bluemke, 2016; Watson et al., 1988).

2.4 Data Analysis

Data Analysis was performed in R using R Version 4.1.0 (R Core Team, 2020) and closely aligned with the analysis of Nussbaum, Schirmer and Schweinberger (2023b; for their published scripts see https://doi.org/10.17605/OSF.IO/3JKCQ). Response omissions (~1 %) were treated as errors and participants with more than 5 % of such omissions were excluded from data analysis. Analyses of Variance (ANOVAs) and correlational analyses were performed on data averaged across speaker and pseudoword. Post-hoc tests were Benjamini-Hochberg corrected where appropriate (Benjamini & Hochberg, 1995).

Concerning the PROMS, I adopted the measure proposed by Nussbaum, Schirmer and Schweinberger (2023b). Responses were coded from 0 to 1 in 0.25 steps starting with the “definitely” correct option down two the “definitely” incorrect option (thus, “don’t know” was always coded with 0.5) and subtracted 0.5 from the final measure. Thus, a positive score indicates that participants were more correct/confident, whereas a negative score indicates more incorrect/uncertain ratings. Then, performance across trials was averaged for each subtest (Nussbaum, Schirmer & Schweinberger, 2023b).

The present study was preregistered; further details regarding the variables and analysis plan can be found in the corresponding OSF registration (https://doi.org/10.17605/OSF.IO/76PV5). Take note that the analyses in the present master thesis deviate from the analysis plan in the preregistration. A subsequent publication will adhere to the preregistered analysis plan.

3. Results

3.1 Demographic, musicality, and personality characteristics of participants

Singers vs. Instrumentalists Singers and instrumentalists did not differ significantly in the socioeconomic status assessed via educational level ($\chi^2(2, N = 88) = 1.06$, $p = .588$), highest academic degree ($\chi^2(7, N = 88) = 9.06$, $p = .249$), and household income ($\chi^2(4, N = 88) = 5.23$, $p = .264$; for more details see Table A.3). Further, the groups did not differ in age or positive and negative affect and were comparable regarding Big Five personality traits and autistic traits. In the Gold-MSI, singers and instrumentalists scored comparatively on the general musicality score, but there were differences on two subfactors: instrumentalists scored higher on the subfactor Formal Education, while singers scored higher on Singing. In the PROMS, both groups performed comparably in all four subtests (see Table 1 for a summary of characteristics assessed via self-report and music performance in the PROMS).

Professionals vs. amateurs vs. non-musicians Professionals, amateurs and non-musicians did not differ in the socioeconomic status assessed via educational level ($\chi^2(6, N = 166) = 11.11$, $p = .085$) and highest academic degree ($\chi^2(16, N = 166) = 24.04$, $p = .089$). However, there were differences regarding household income ($\chi^2(8, N = 166) = 20.19$, $p = .01$, $\phi = .25$), with amateurs reporting higher household income than professionals and non-musicians (for more details see Table A.4). The groups were comparable in age as well as in positive and negative affect. For the Big Five, analyses of variance revealed group differences for extraversion, with slightly higher levels in professionals than in amateurs ($|t(82.15)| = 2.91$, $p = .005$, $d = 0.64$ [0.20, 1.08]; see Table 2 for a summary of characteristics assessed via self-report).

Regarding autistic traits, the three groups did not differ in their overall score, but there were differences on the two subscales proposed by Hoekstra et al. (2008). Non-musicians scored lower on the *Attention to Details* subscale compared to professionals ($|t(75.87)| = 2.42$, $p = .018$, $d = 0.56$ [0.09, 1.01]) and amateurs ($|t(83.53)| = 2.87$, $p = .005$, $d = 0.63$ [0.19, 1.07]). On the *Social* subscale, professionals scored lower compared to amateurs ($|t(101.75)| = 2.90$, $p = .005$, $d = 0.57$ [0.18, 0.97]) and non-musicians ($|t(67.08)| = 2.32$, $p = .024$, $d = 0.57$ [0.08, 1.05]). When split into the five subscales as originally proposed by Baron-Cohen et al. (2001), there were group differences on the *Social Skills* subscale, with professionals scoring lower than amateurs and non-musicians (professionals vs. amateurs: $|t(107.81)| = 3.36$, $p = .001$, $d = 0.65$ [0.26, 1.03]; professionals vs. non-musicians: $|t(62.40)| = 2.25$, $p = .028$, $d = 0.57$ [0.06, 1.07]; see also Table 2; for all post-hoc tests refer to Tables C.1-C.3).

In the Gold-MSI, professionals scored significantly higher than non-musicians on all subfactors and the general musicality score (all |*t*s| ≥ 3.79, *p*s < .001) and higher than amateurs on all subfactors as well as the general musicality score (all |*t*s| ≥ 4.08, *p*s < .001), except Emotion (|*t*(80.76)|= 2.29, *p* = .025). Amateurs scored considerably higher than non-musicians on the general musicality score and all subfactors (all |*t*s| ≥ 2.59 , *p*s < .013; for details see Table 2; for all post-hoc tests refer to Tables C.4-C.6).

In the PROMS, professionals outperformed non-musicians in all subtests (all |*t*s| ≥ 2.99, *p*s ≤ .004). Professionals also performed better than amateurs in the Pitch and Melody subtest (Pitch: |*t*(97.32)| = 2.57, *p* = .012, *d* = 0.55 [0.12, 0.98]; Melody: |*t*(95.24)| = 4.42, *p* < .001, *d* = 0.91 [0.48, 1.33]), whereas there were no differences in the Timbre and Rhythm subtests (*p*s ≥ .09). Amateurs performed better than non-musicians in the Pitch, Melody and Rhythm subtest (Pitch: |*t*(81.21)| = 4.39, *p* < .001, *d* = 0.97 [0.51, 1.43]; Melody: |*t*(91.34)| = 5.65, *p* < .001, *d* = 1.18 [0.74, 1.62]; Rhythm: |*t*(80.84)| = 3.16, *p* = .002, *d* = 0.7 [0.25, 1.15]), but not in the Timbre subtest (*p* = .064; for all analyses and post-hoc tests refer to Supplemental Analyses (C.7) and Tables C.8-C.10).

Table 1

Characteristics of participants (singers vs. instrumentalists) - Demography, personality, and musicality

	Singers	**Instrumentalists**					
	M (SD)	*M (SD)*	*t*	*df*[a]	*p*	*Cohens d*	
Age	27.02 (8.2)	28.51 (10.6)	-0.73	78.93	.465	-0.17 [-0.61, 0.28]	
PANAS							
positive Affect	3.00 (0.68)	3.11 (0.57)	-0.77	84.78	.446	-0.17 [-0.59, 0.26]	
negative Affect	1.53 (0.47)	1.40 (0.35)	1.49	80.61	.141	0.33 [-0.11, 0.77]	
Big Five							
Openness	4.04 (0.55)	3.99 (0.51)	0.46	85.96	.647	0.10 [-0.32, 0.52]	
Conscientiousness	3.47 (0.69)	3.76 (0.70)	-1.91	85.62	.060	-0.41 [-0.84, 0.02]	
Extraversion	3.21 (0.70)	3.00 (0.73)	1.44	85.3	.155	0.31 [-0.12, 0.74]	
Agreeableness	3.81 (0.57)	4.01 (0.60)	-1.61	85.29	.112	-0.35 [-0.77, 0.08]	
Neuroticism	2.74 (0.77)	2.61 (0.78)	0.80	85.75	.426	0.17 [-0.25, 0.60]	
AQ							
Total	18.2 (6.15)	19.28 (8.55)	-0.68	76.04	.500	-0.16 [-0.60, 0.30]	
Attention to Detail	5.4 (2.33)	5.63 (2.53)	-0.44	84.64	.662	-0.10 [-0.52, 0.33]	
Social	12.8 (5.37)	13.65 (7.53)	-0.61	75.65	.545	-0.14 [-0.59, 0.31]	
Social Skills	2.4 (1.94)	3.09 (2.95)	-1.30	72.01	.200	-0.31 [-0.77, 0.16]	
Communication	2.53 (1.94)	2.44 (2.3)	0.20	82.02	.842	0.04 [-0.39, 0.48]	
Imagination	2.51 (1.75)	2.81 (1.88)	-0.78	84.86	.437	-0.17 [-0.60, 0.26]	
Attention Switching	5.36 (1.91)	5.30 (2.23)	0.12	82.69	.904	0.03 [-0.40, 0.46]	
Gold-MSI							
General ME	4.78 (0.85)	4.75 (0.80)	0.19	85.99	.866	0.04 [-0.39, 0.46]	
Active Engagement	3.83 (0.82)	4.21 (1.13)	-1.79	76.79	.078	-0.41 [-0.86, 0.05]	
Formal Education	**4.39 (1.14)**	**4.95 (0.62)**	-2.85	68.31	.006	-0.69 [-1.18, -0.20]	**
Emotion	5.5 (0.81)	5.60 (0.76)	-0.60	85.99	.549	-0.13 [-0.55, 0.29]	
Singing	**4.98 (0.97)**	**4.19 (1.27)**	3.25	78.56	.002	0.73 [0.27, 1.19]	**
Perception	5.73 (0.82)	5.77 (1.03)	-0.22	80.11	.825	-0.05 [-0.49, 0.39]	
PROMS							
Pitch	0.23 (0.08)	0.24 (0.06)	-0.30	82.92	.766	-0.07 [-0.50, 0.37]	
Melody	0.17 (0.1)	0.14 (0.10)	1.29	85.08	.199	0.28 [-0.15, 0.71]	
Timbre	0.29 (0.08)	0.3 (0.09)	-0.59	86.00	.556	-0.13 [-0.55, 0.30]	
Rhythm	0.31 (0.09)	0.30 (0.08)	-0.56	85.13	.577	-0.12 [-0.55, 0.30]	

Note. Descriptive values show mean ratings for the PANAS (Breyer & Bluemke, 2016), the Big Five Domains (Rammstedt et al., 2020), and the Gold-MSI (Müllensiefen et al., 2014). AQ scores were calculated based on Hoekstra et al. (2008) and Baron-Cohen et al. (2001).

[a] Note that original degrees of freedom were 86 but were corrected due to unequal variance.

Table 2

Characteristics of participants (professionals vs. amateurs vs. non-musicians) - Demography, personality, and musicality

	Pro-fessionals	Amateurs	Non-Musicians				
	M (SD)	*M (SD)*	*M (SD)*	F^a	*p*	ω^2	
Age	29.55 (5.58)	27.75 (9.44)	30.5 (6.54)	1.77	.173	0.01[0.00, 0.05]	
PANAS							
positive Affect	3.32 (0.65)	3.05 (0.63)	3.1 (0.67)	2.34	.099	0.02 [0.00, 0.07]	
negative Affect	1.69 (0.48)	1.47 (0.42)	1.49 (0.69)	2.82	.063	0.02 [0.00, 0.08]	
Big Five							
Openness	4.12 (0.50)	4.02 [0.53)	3.81 (0.80)	2.91	.057	0.02 [0.00, 0.08]	
Conscientiousness	3.49 (0.71)	3.61 (0.70)	3.76 (0.72)	1.43	.242	0.005 [0.0, 0.04]	
Extraversion	**3.48 (0.66)**	**3.11 (0.72)**	3.38 (0.79)	4.39	.014	0.04 [0.00, 0.11]	*
Agreeableness	3.92 (0.57)	3.91 (0.59)	3.75 (0.66)	1.14	.324	0.002 [0.0, 0.02]	
Neuroticism	2.95 (0.65)	2.69 (0.77)	2.58 (0.82)	2.65	.074	0.02 [0.00, 0.07]	
AQ							
Total	15.7 (4.98)	18.73 (7.40)	17.58 (6.41)	2.85	.061	0.02 [0.00, 0.08]	
Attention to Detail	**5.43 (2.04)**	**5.51 (2.42)**	**4.32 (2.01)**	4.00	.020	0.03 [0.00, 0.10]	*
Social	**10.28 (4.70)**	**13.22 (6.49)**	**13.26 (6.51)**	3.55	.031	0.03 [0.00, 0.09]	*
Social Skills	**1.48 (1.68)**	**2.74 (2.49)**	**2.61 (2.63)**	4.13	.018	0.04 [0.00, 0.10]	*
Communication	1.85 (1.61)	2.49 (2.12)	2.39 (1.73)	1.56	.213	0.007 [0.0, 0.04]	
Imagination	2.18 (1.52)	2.66 (1.8)	2.87 (1.95)	1.62	.201	0.01 [0.00, 0.04]	
Attention Switching	4.78 (1.91)	5.33 (2.06)	5.39 (1.92)	1.27	.284	0.003 [0.0, 0.03]	
Gold-MSI							
General ME	**5.68 (0.50)**	**4.76 (0.82)**	**2.74 (1.07)**	132.0	<.001	0.61 [0.52, 0.68]	***
Active Engagement	**4.94 (0.81)**	**4.02 (1.00)**	**2.95 (1.19)**	38.02	<.001	0.31 [0.19, 0.41]	***
Formal Education	**5.95 (0.56)**	**4.66 (0.96)**	**1.71 (0.68)**	280.5	<.001	0.77 [0.71, 0.81]	***
Emotion	**5.88 (0.73)**	**5.55 (0.78)**	**4.95 (1.32)**	10.11	<.001	0.10 [0.02, 0.19]	***
Singing	**5.34 (0.83)**	**4.59 (1.19)**	**2.84 (1.26)**	51.23	<.001	0.38 [0.26, 0.47]	***
Perception	**6.31 (0.51)**	**5.75 (0.92)**	**4.22 (1.49)**	45.81	<.001	0.35 [0.24, 0.45]	***
PROMS							
Pitch	**0.27 (0.06)**	**0.24 (0.07)**	**0.18 (0.06)**	16.87	<.001	0.16 [0.07, 0.26]	***
Melody	**0.23 (0.08)**	**0.16 (0.10)**	**0.07 (0.08)**	32.56	<.001	0.28 [0.16, 0.38]	***
Timbre	**0.32 (0.08)**	**0.29 (0.08)**	**0.26 (0.09)**	4.89	.009	0.04 [0.00, 0.11]	**
Rhythm	**0.33 (0.08)**	**0.32 (0.09)**	**0.27 (0.08)**	6.23	.002	0.06 [0.01, 0.14]	**

Note. Descriptive values show mean ratings for the PANAS (Breyer & Bluemke, 2016), the Big Five Domains (Rammstedt et al., 2020), and the Gold-MSI (Müllensiefen et al., 2014). AQ scores were calculated based on Hoekstra et al. (2008) and Baron-Cohen et al. (2001).

[a] Degrees of freedom for all tests were F(2, 163).

3.2 Emotion classification performance

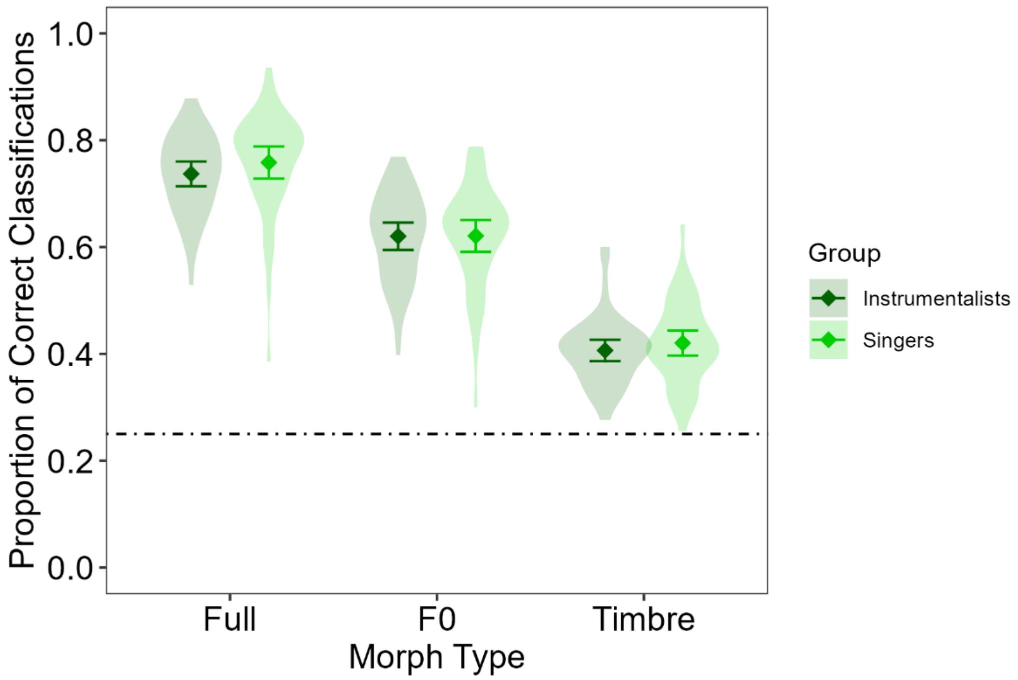
3.2.1 Singers vs. instrumentalists

The mean proportion of correct responses was submitted to a mixed ANOVA with Emotion (Happiness, Pleasure, Fear and Sadness) and Morph Type (Full, F0 and Timbre) as repeated measures factors and Group (singers and instrumentalists) as a between subject factor. Reference stimuli (emotional averages) were excluded from this analysis.

The results included main effects of **Emotion** ($F(3, 258) = 72.43$, $p < .001$, $\omega^2 = .45$, 95%-CI [0.36, 0.52]) and **Morph Type** ($F(2, 172) = 768.93$, $p < .001$, $\omega^2 = .90$, 95%-CI [0.87, 0.92], $\varepsilon_{HF} = .741$). There was no significant main effect of **Group** ($F(1, 86) = 0.38$, $p = .542$). The significant main effects were qualified by a significant interaction of **Emotion × Morph Type** ($F(6, 516) = 22.78$, $p < .001$, $\omega^2 = .20$, 95%-CI [0.14, 0.25], $\varepsilon_{HF} = .827$). The three-way interaction did not reach significance ($F(6, 516) = 1.33$, $p = .249$).

Figure 3

Mean proportion of correct responses per Morph Type separately for singers and instrumentalists



Note. Whiskers represent 95 % confidence intervals. Violin plots represent variation of individual participants. The dotted line represents guessing rate at .25.

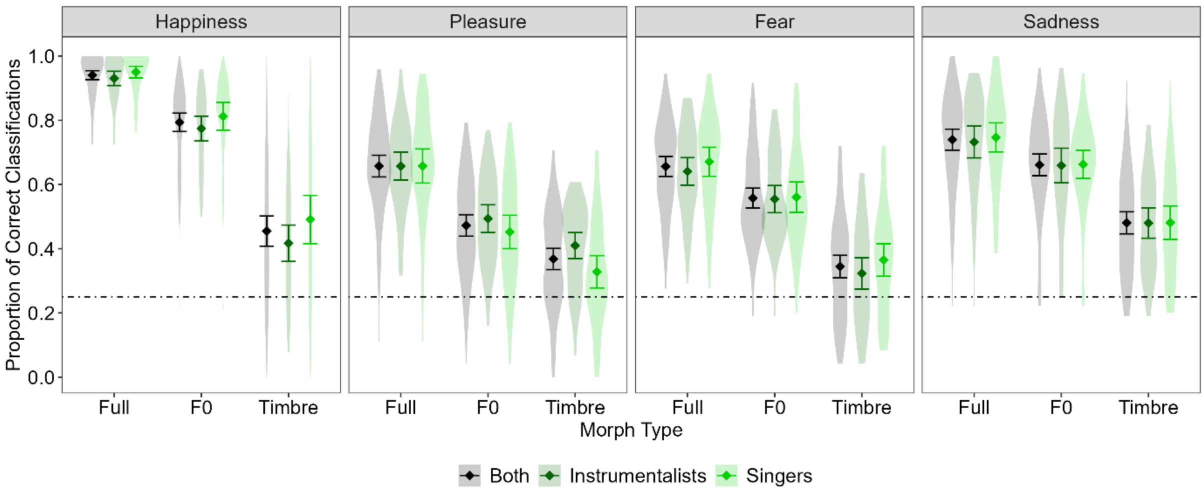
Follow-up analyses of the Morph Type effect revealed that performance was best in the Full condition ($M = 0.75 \pm 0.02$ SEM), followed by the F0 ($M = 0.62 \pm 0.02$) and then the Timbre condition ($M = 0.41 \pm 0.01$; Full vs. F0: $|t(87)| = 23.28$, $p < .001$, $d = 2.50$ [2.07, 2.92], F0 vs Timbre: $|t(87)| = 21.44$, $p < .001$, $d = 2.30$ [1.90, 2.70], Full vs Timbre: $|t(87)| = 33.94$, $p < .001$, $d = 3.64$ [3.06, 4.21]). This main effect of Morph Type was also found for all emotions separately (all $Fs(2, 174) > 102.44$, $p < .001$), although it differed slightly between emotions, as suggested by the interaction (see Figure 4).

To address the specific interest in the relative importance of F0 and Timbre for the different emotions, we calculated the performance difference$_{\text{F0-Tbr}}$ for each emotion separately. Performance difference was largest for Happiness ($M = 0.34 \pm 0.02$ SEM), followed by Fear ($M = 0.21 \pm 0.02$), Sadness ($M = 0.18 \pm 0.02$), and Pleasure ($M = 0.10 \pm 0.02$; with pairwise comparisons $|ts(87)| \geq 2.57$, $ps_{\text{corrected}} < .012$, $ds \geq 0.28$ [0.06, 0.49], except for Fear vs. Sadness ($|t(87)| = 1.13$, $p_{\text{corrected}} = .261$).

With the exception of the non-significant main effect of Group, these results are a replication of the results found by Nussbaum, Schirmer and Schweinberger (2023b).

Figure 4

Mean proportion of correct responses per Emotion and Morph Type for singers and instrumentalists



Note. Whiskers represent 95 % confidence intervals. Violin plots represent variation of individual participants. The dotted line represents guessing rate at .25.

3.2.2 Professionals vs. amateurs vs. non-musicians

The mean proportion of correct responses was submitted to a mixed ANOVA with Emotion (Happiness, Pleasure, Fear and Sadness) and Morph Type (Full, F0 and Timbre) as repeated measures factors and Group (professionals, amateurs, non-musicians) as a between subject factor. Reference stimuli (emotional averages) were excluded from this analysis.

The results included main effects of **Emotion** ($F(3, 489) = 130.24$, $p < .001$, $\omega^2 = .44$, 95%-CI [0.38, 0.49]) and **Morph Type** ($F(2, 326) = 1357.80$, $p < .001$, $\omega^2 = .89$, 95%-CI [0.87, 0.91], $\varepsilon_{HF} = .829$). There was no significant main effect of **Group** ($F(2, 163) = 1.96$, $p = .144$). The significant main effects were qualified by a significant interaction of **Emotion × Morph Type** ($F(6, 978) = 40.95$, $p < .001$, $\omega^2 = .20$, 95%-CI [0.15, 0.24], $\varepsilon_{HF} = .865$). The interaction of **Group x Morph Type** was not significant, but showed a trend ($F(4, 326) = 2.14$, $p = .089$). The three-way interaction did not reach significance ($F(12, 978) = 0.74$, $p = .688$).

In exploration of the trend of the Group x Morph Type interaction, post-hoc tests revealed that professionals outperformed non-musicians in Full- and F0-morph conditions, whereas there was no difference in the Timbre-morph condition (Full: $|t(73.34)| = 2.88$, $p = .005$, $d = 0.67$ [0.20, 1.14]; F0: $|t(67.52)| = 2.38$, $p = .020$, $d = 0.58$ [0.09, 1.06]; Timbre: $|t(76.00)| = 0.21$, $p = .832$, $d = 0.05$ [−0.40, 0.50]). There were no differences in performance between professionals and amateurs in any morph condition (all $|ts| \leq 0.99$, $ps \geq .322$). Amateurs outperformed non-musicians in the Full-morph condition, but not in the F0- or Timbre-morph condition (Full: $|t(84.38)| = 2.04$, $p = .044$, $d = 0.44$ [0.01, 0.88]; F0: $|t(83.72)| = 1.64$, $p = .105$, $d = 0.36$ [-0.08, 0.79]; Timbre: $|t(86.74)| = 0.03$, $p = .975$, $d = 0.007$ [-0.41, 0.43]; see Figure 5; for all post-hoc tests refer to Tables C.11-C.13).

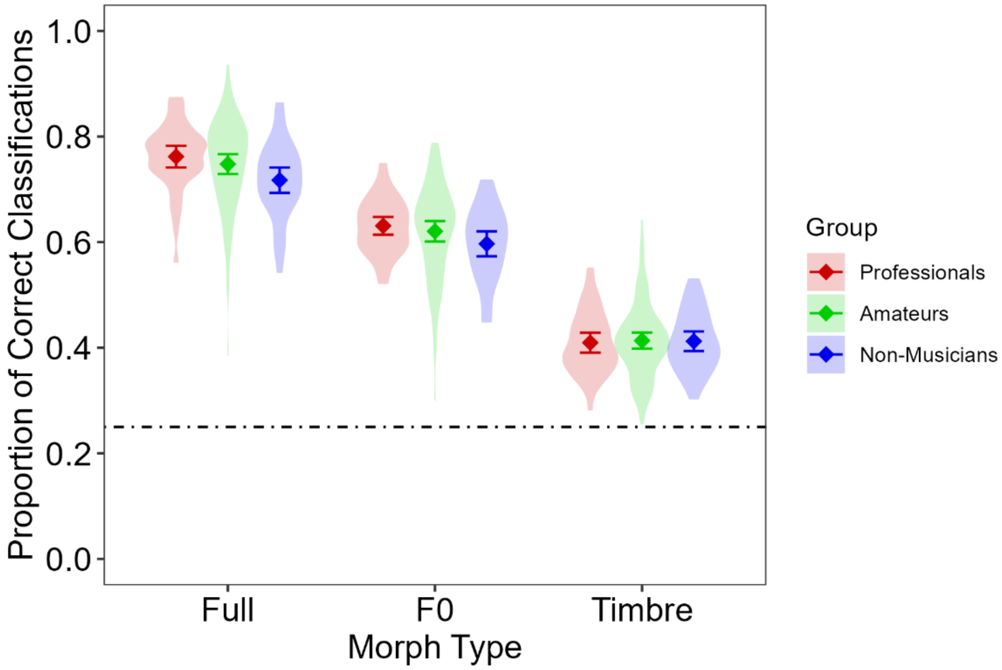
Follow-up analyses of the Morph Type effect revealed that performance was best in the Full-morph condition ($M = 0.74 \pm 0.01$ SEM), followed by the F0- ($M = 0.62 \pm 0.01$) and then the Timbre-morph condition ($M = 0.41 \pm 0.004$; Full vs. F0: $|t(165)| = 29.13$, $p < .001$, $d = 2.27$ [1.98, 2.56]; F0 vs Timbre: $|t(165)| = 31.00$, $p < .001$, $d = 2.41$ [2.11, 2.71]; Full vs Timbre: $|t(165)| = 49.60$, $p < .001$, $d = 3.86$ [3.41, 4.30]). This main effect of Morph Type was also found for all emotions separately (all F s(2, 330) > 210.63, $p < .001$), although it differed slightly between emotions, as suggested by the interaction.

Again, the performance difference$_{\text{F0-Tbr}}$ was calculated for each emotion separately. Performance difference was largest for Happiness ($M = 0.33 \pm 0.03$ SEM), followed by Fear

(M = 0.22 ± 0.02), Sadness (M = 0.17 ± 0.03), and Pleasure (M = 0.10 ± 0.02; with all pairwise comparisons $|ts(165)| \geq 2.54$, $ps_{corrected} < .012$, $ds \geq 0.20$ [0.04, 0.35]).

Figure 5

Mean proportion of correct responses per Morph Type separately for professionals, amateurs and non-musicians



Note. Whiskers represent 95 % confidence intervals. Violin plots represent variation of individual participants. The dotted line represents guessing rate at .25.

3.3 Links between musical skills and vocal emotion perception

In a subsequent exploratory analysis, we calculated Spearman correlations between vocal emotion perception performance and both the PROMS music perception performance and the Gold-MSI self-rated musicality. The results are shown in Figure 6 as well as Tables 3 and 4.

3.3.1 Correlations between the PROMS and vocal emotion perception

Singers vs. instrumentalists In the group of amateurs, a strong correlation between the overall vocal emotion recognition performance and average PROMS performance was found ($\rho(86) = .39$, $p = .002$). This correlation also emerged in a separate analysis of singers ($\rho(43) = .52$, $p = .005$), but was non-significant in instrumentalists ($\rho(41) = .23$, $p = .326$), possibly due to reduced variance. Performance in the Full- and F0-morph conditions correlated with all subtests of the PROMS except the pitch subtest. There was no link between the Timbre-morph condition and the timbre subtest, but a correlation with the rhythm subtest.

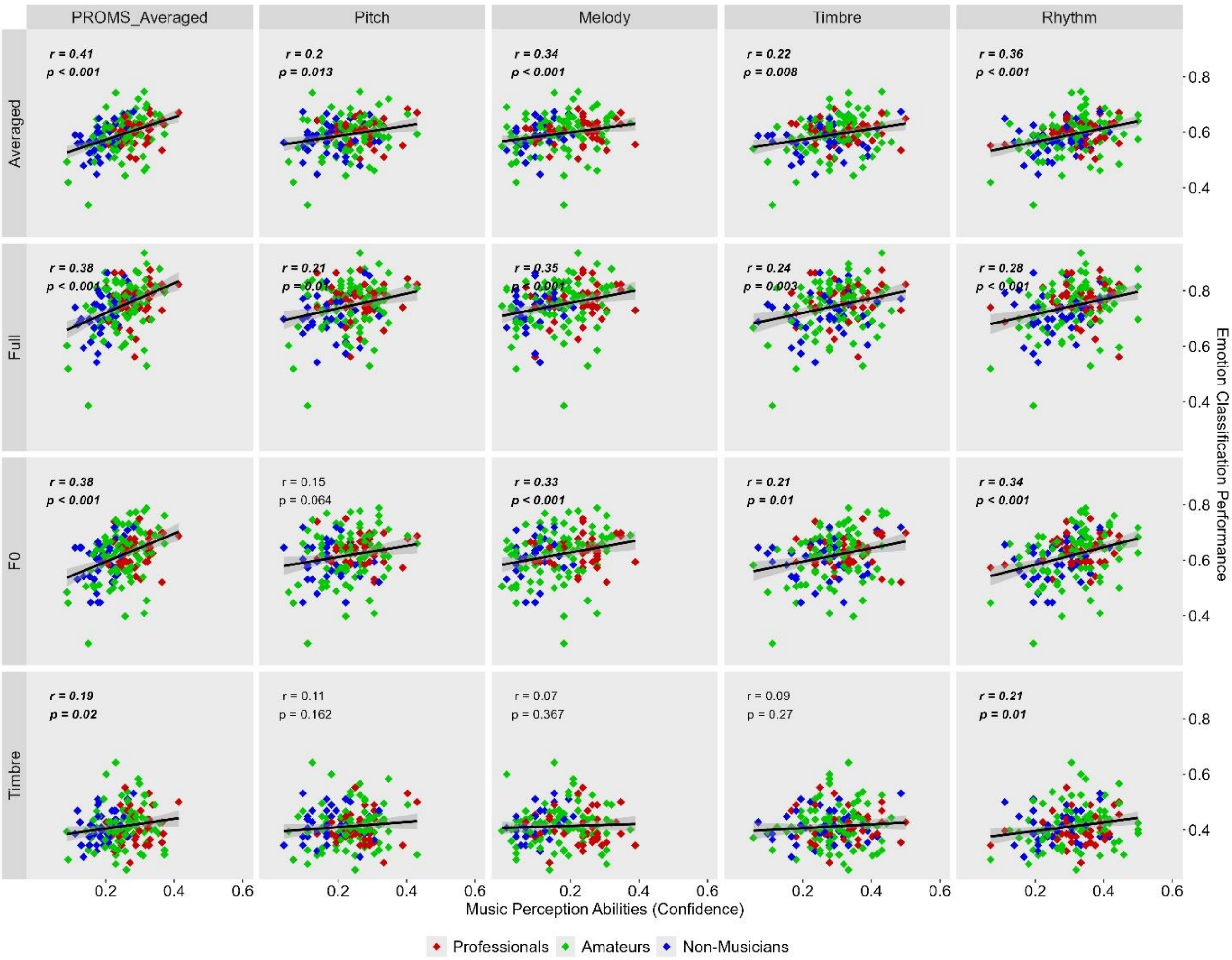
In the next step, we explored these correlations in more detail to examine a potential role of musical training. Specifically, we calculated partial correlations to control for musical training (Kim, 2015). The correlations between VER_{Avg} and $PROMS_{Avg}$ ($\rho(86) = .38$, $p = .003$), Full-Morphs and Melody ($\rho(86) = .28$, $p = .023$), Full-Morphs and Rhythm ($\rho(86) = .32$, $p = .008$), F0-Morphs and Melody ($\rho(86) = .34$, $p = .006$), F0-Morphs and Rhythm ($\rho(86) = .32$, $p = .008$) remained significant. Correlations of F0-Morphs and Timbre ($\rho(86) = .24$, $p = .053$) and of Timbre-Morphs with Rhythm ($\rho(86) = .22$, $p = .062$) turned marginally non-significant when controlling for musical training.

Professionals vs. amateurs vs. non-musicians Analysis for all three groups yielded a strong correlation between the overall vocal emotion recognition performance and average PROMS performance ($\rho(164)= .41$, $p < .001$). This correlation also emerged in a separate analysis of amateurs ($\rho(86) = .39$, $p = .002$) and non-musicians ($\rho(36) = .48$, $p = .029$), but was non-significant in professionals ($\rho(38) = .21$, $p = .193$), possibly due to reduced variance. Performance in the Full-morph condition correlated with all subtests of the PROMS. Performance in the F0-morph condition correlated with all subtests of the PROMS, except the Pitch subtest. There was no link between the Timbre-morph condition and the Timbre subtest, but a correlation with the Rhythm subtest.

Similarly to the procedure with the amateurs, these correlations were explored in more detail to examine a potential role of musical training and partial correlations to control for musical training were calculated (Kim, 2015). The correlations between VER_{Avg} and $PROMS_{Avg}$ ($\rho(164) = .39$, $p < .001$), Full-Morphs and Melody ($\rho(164) = .32$, $p < .001$), Full-Morphs and Rhythm ($\rho(164) = .25$, $p = .003$), Full-Morphs and Timbre ($\rho(164) = .21$, $p = .012$), F0-Morphs and Melody ($\rho(164) = .31$, $p < .001$), F0-Morphs and Rhythm ($\rho(164) = .32$, $p < .001$), F0-Morphs and Timbre ($\rho(164) = .19$, $p = .021$), and Timbre-Morphs and Rhythm ($\rho(164) = .19$, $p = .021$) remained significant. The correlation of Full-Morphs with Pitch turned non-significant when controlling for musical training ($\rho(164) = .15$, $p = .066$).

Figure 6

Correlations between vocal emotion recognition and music perception performance (PROMS) for professionals, amateurs and non-musicians



Note. The x-axis shows the different subtests of the PROMS (Pitch, Melody, Timbre, and Rhythm) as well as the averaged performance across all subtests (PROMS_Averaged). The y-axis shows the vocal emotion classification performance separately for each Morph Type (Full, F0 and Timbre) and averaged across Morph Types (Averaged). p-values were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995; false discovery rate set to 0.05, total number of tests = 20).

Table 3

Correlations between vocal emotion recognition and music perception performance (PROMS)

	$PROMS_{Avg}$	Pitch	Melody	Timbre	Rhythm
Amateurs (N = 88)					
VER_{Avg}	**.39 (.002)**	.17 (.142)	**.29 (.017)**	.23 (.054)	**.38 (.002)**
Full-Morphs	**.34 (.005)**	.14 (.219)	**.27 (.022)**	**.23 (.049)**	**.31 (.009)**
F0-Morphs	**.39 (.002)**	.16 (.186)	**.34 (.005)**	**.24 (.044)**	**.32 (.008)**
Timbre-Morphs	.22 (.054)	.12 (.305)	.08 (.473)	.10 (.352)	**.25 (.039)**
All groups (N = 166)					
VER_{Avg}	**.41 (<.001)**	**.20 (.013)**	**.34 (<.001)**	**.22 (.008)**	**.36 (<.001)**
Full-Morphs	**.38 (<.001)**	**.21 (.01)**	**.35 (<.001)**	**.24 (.003)**	**.28 (<.001)**
F0-Morphs	**.38 (<.001)**	.15 (.064)	**.33 (<.001)**	**.21 (.010)**	**.34 (<.001)**
Timbre-Morphs	**.19 (.020)**	.11 (.162)	.07 (.367)	.09 (.270)	**.21 (.010)**

Note. VER = Vocal Emotion Recognition performance. p-values (in parenthesis) were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995).

3.3.2 Correlations between the Gold-MSI and vocal emotion perception

Singers vs. instrumentalists There was no correlation between vocal emotion perception performance and self-rated general musical sophistication ($\rho(86) = .12$, $p = .256$). Separate analysis for singers and instrumentalists yielded the same results (singers: $\rho(43) = .28$, $p = .064$; instrumentalists: $\rho(41) = -.09$, $p = .574$).

Professionals vs. amateurs vs. non-musicians There was a correlation between vocal emotion perception performance and self-rated general musical sophistication ($\rho(164) = .2$, $p = .049$), which turned non-significant when controlling for musical training ($\rho(164) = .12$, $p = .345$). Further, self-rated singing abilities were marginally linked to increased sensitivity towards vocal emotions ($\rho(164) = .21$, $p = .049$), but not when controlled for musical training ($\rho(164) = .15$, $p = .261$).

All partial correlations as well as exploratory correlations between the Gold-MSI and the PROMS can be found in the Supplemental Analyses (Tables C.14-C.16).

Table 4

Correlations between vocal emotion recognition and self-rated musicality (Gold-MSI)

	General Musicality	Active Engagement	Formal Education	Emotion	Singing	Perception
Amateurs (N = 88)						
VER_{Avg}	.12 (.256)	-.1 (.355)	.10 (.366)	.21 (.051)	.12 (.285)	.10 (.336)
Full-Morphs	.03 (.748)	-.12 (.266)	.01 (.933)	.12 (.260)	.07 (.491)	.03 (.816)
F0-Morphs	.15 (.163)	-.03 (.763)	.04 (.730)	.17 (.112)	.15 (.153)	.14 (.188)
Timbre-Morphs	.04 (.705)	-.11 (.319)	.17 (.115)	.12 (.196)	.00 (.983)	.00 (.970)
All groups (N = 166)						
VER_{Avg}	**.20 (.049)**	.04 (.689)	.16 (.090)	.11 (.296)	**.21 (.049)**	.19 (.060)
Full-Morphs	**.21 (.049)**	.06 (.595)	.17 (.076)	.10 (.311)	**.24 (.049)**	.18 (.063)
F0-Morphs	.18 (.063)	.05 (.641)	.11 (.297)	.07 (.517)	**.20 (.049)**	.17 (.078)
Timbre-Morphs	.04 (.649)	-.02 (.842)	.08 (.445)	.05 (.641)	.03 (.748)	.05 (.641)

Note. VER = Vocal Emotion Recognition performance. p-values (in parenthesis) were adjusted for multiple comparisons using the Benjamini-Hochberg correction when appropriate (Benjamini & Hochberg, 1995).

3.3.3 Correlations between personality traits and vocal emotion perception

To rule out that any performance difference could be attributed to one of the personality traits that differed between groups, a correlation with the averaged vocal emotion performance was executed.

Singers vs. instrumentalists The results were non-significant for all subscales of the Big Five Inventory (all *p*s $\geq$ .203). None of the AQ scales correlated significantly with vocal emotion perception performance (all $ps_{corrected} \geq .389$).

Professionals vs. amateurs vs. non-musicians The results were non-significant for all subscales of the Big Five Inventory (all *p*s $\geq$ .074). None of the AQ scales correlated significantly with vocal emotion perception performance (all $ps_{corrected} \geq .795$).

4. Discussion

In the present study, we compared the vocal emotion perception performance of non-professional singers and instrumentalists using stimuli which were acoustically manipulated via parameter-specific voice-morphing. Further, we targeted specifically non-professional musicians (amateurs) in contrast to professional musicians and non-musicians.

As predicted, singers and instrumentalists did not significantly differ in their vocal emotion perception performance, neither in a condition with full emotion modulation nor in conditions where only pitch (F0) or timbre were emotionally informative. The comparison of professionals, amateurs and non-musicians revealed that amateurs performed equally to professionals in full as well as F0- and timbre-modulated conditions. Further, they performed better than non-musicians in a condition with full emotion modulation, but not in a condition where only pitch was emotionally informative. In both these conditions, professionals outperformed non-musicians. There were no differences between the groups regarding vocal emotion recognition performance in a condition where only timbre was emotionally informative.

In the following, these results will be discussed in more detail.

4.1 Singers vs. instrumentalists

As expected, vocal emotion recognition performance of singers and instrumentalists did not differ in any condition. This corroborates previous findings (Francisco, 2021) and reflects neuroimaging evidence that vocal emotions are processed similarly among singers and instrumentalists (I. Martins et al., 2022). In the light of amateurs outperforming non-musicians in overall vocal emotion perception, these results lend strength to the idea that musicality in general benefits vocal emotion perception, with no significant differences between singing vs. playing an instrument.

Instrumentalists reported to have slightly more formal education than singers, probably due to the fact that mastering an instrument to the point where one can play in an orchestra arguably takes more training than being able to sing in a choir. The initial threshold for playing an instrument is simply much higher than for singing, starting with the acquisition of an instrument and the necessary fine-motoric skills to translate musicality into, for example, hand movements. In most cases, playing in an orchestra is preceded by structured music lessons, which is not strictly necessary for choir singing at an amateur level. This was also reflected in the fact that some singers reported in debriefing to only have sung in a choir for a couple of months with no prior musical practice or lessons. Counterbalancing that, other singers had been active for several decades. Instrumentalists were located mainly in the middle of the spectrum. In this line, the present study provides further evidence that the amount of music training is not a major influence on vocal emotion perception (Schellenberg & Lima, 2024). Moreover, correlational analyses revealed a strong link between music perception skills and vocal emotion perception in the group of amateurs that remained when controlling for music training. This correlation was particularly apparent in a separate analysis of the singers, presumably due to a higher variance in comparison to instrumentalists, which mirrors the higher variance concerning musical experience.

Based on evidence that different instrumentalists have processing advantages for their respective instrument (Margulis et al., 2009; Pantev et al., 2001; Shahin et al., 2008), it is conceivable that singers are particularly tuned to vocal timbre. While we cannot make conclusions about possible differences regarding musical vocal timbre, the present study did not find any differences between singers and instrumentalists concerning timbre-dependent emotion perception in speech. As we only implemented behavioral measures and did not examine the neural responses of participants, the possibility of underlying processing differences cannot be ruled out (e.g., Nikjeh et al., 2008; Tervaniemi, 2009). In the same vein, instrument-specific processing advantages do not show up reliably in behavioral data (Holmes et al., 2022; Weiss, Vanzella, et al., 2015). However, the voice is a unique case of musical timbre as it is not primarily musical. Due to its biological significance, the voice may be far more familiar and/or may attract more attention than any instrument (Belin et al., 2000; Blasi et al., 2011), resulting in an evolutionary advantage for processing vocal cues over instrumental cues regardless of instrumental expertise. Therefore, while instrumentalists may have instrument-specific processing advantages, those could be outweighed by processing advantages for the human voice, leading to vocal prosodic cues being processed similarly by instrumentalists and singers. In this line, previous research found vocal melodies to be better remembered than instrumental melodies both by pianists and non-pianists (Weiss, Vanzella, et al., 2015). Moreover, this effect was found more generally in adults and children (Weiss et al., 2012; Weiss, Schellenberg, et al., 2015). Additionally, musical emotions were found to be represented by similar activation patterns in the auditory cortex as vocal emotions of the human voice (Sachs et al., 2018).

4.2 Professionals vs. amateurs vs. non-musicians

There was no significant main effect of group when comparing professional musicians, amateurs and non-musicians. As we defined the groups based on formal music education (professionals), current musical engagement (amateurs) and meeting neither of those criteria (non-musicians), it can be concluded that the relevant difference regarding participants' vocal emotion perception was not whether they had music training or engaged in musical activities. Instead, it is likely that participants differed on where they were located on a spectrum of auditory sensitivity. Exploratory analyses on correlational patterns based on the PROMS showed a strong positive association of music perception skills and vocal emotion recognition performance across the whole sample, crucially even when controlling for music training. Thus, auditory perception skills rather than music training mediated the link between musicality and vocal emotion perception, which corroborates previous findings (Correia et al., 2022; Nussbaum, Schirmer & Schweinberger, 2023b). This notion is further supported by findings that musically untrained individuals with higher music perception skills showed enhanced neural encoding and perception of speech (Mankel & Bidelman, 2018; Swaminathan & Schellenberg, 2017).

As expected, amateurs and professional musicians did not differ in their vocal emotion perception performance, neither in full emotion modulation nor in conditions with either F0- or timbre-modulated emotions, similar to singers vs. instrumentalists. While there is evidence for qualitative differences between amateurs and professionals concerning neural activation patterns during musical performance (Kleber et al., 2010; Lotze et al., 2003), those do not seem to extend to behavioral vocal emotion perception performance. Assuming that a main difference between amateurs and professional musicians regarding their musical expertise lies in professionals undergoing more extensive formal training and/or practicing more regularly (as seen in the corresponding Gold-MSI scores), the similarities in performance further support the hypothesis that the amount of (formal) music training is not the main influence of musicality on vocal emotion perception.

In planned post-hoc exploration of the interaction of group and morph type we found group differences for full and F0-modulated emotion perception. While there were no significant differences between the vocal emotion perception performances of amateurs and professionals, there were slight differences when comparing the two musician groups to the group of non-musicians. Amateurs were better than non-musicians only in the condition with full emotion modulation, whereas professionals outperformed non-musicians additionally in the condition where only pitch was emotionally informative. This may be a sign of different levels of auditory sensitivity that were also observed in the music perception skills of the three groups: professionals were better than amateurs in the PROMS subtests concerning pitch and melody discrimination, the differences having medium and large effect sizes respectively, and both musician groups outperformed non-musicians. While these differences between professionals and amateurs were not pronounced enough to distinguish their overall vocal emotion perception performances, the advantage in (musical) pitch and melody perception skills professionals had over amateurs may have been enough to give them an edge when comparing the two groups to non-musicians in a condition where emotion recognition was solely dependent on pitch perception.

This further illustrates that auditory sensitivity concerning speech perception especially involves a superior ability to perceive emotional pitch cues (Nussbaum & Schweinberger, 2021; Strait et al., 2009). This was not the case for timbre cues, as there were no differences between the groups when only timbre was emotionally informative. In that context, it is also relevant that individuals with higher musical abilities seem to rely more on pitch than on timbre cues for prosody perception (Cui & Kuang, 2019). Additionally, the emotion recognition accuracy was higher in F0-modulated conditions than in timbre-modulated conditions, across all emotional categories and groups, which implies pitch to be more important for emotion recognition than timbre (Nussbaum, Schirmer & Schweinberger, 2023b). Thus, our findings highlight that auditory sensitive individuals excel at using pitch which is considered the more dominant acoustic cue for recognizing vocal emotions.

The importance of pitch is further corroborated by exploratory correlational patterns, which also highlight a particular importance of the dynamic pitch contour. Examining the correlational patterns based on the different subtests of the PROMS revealed that the Melody and Rhythm subtests were strong in predicting emotion perception performance (with explained variances between 11.6 and 13 %), while the subtests regarding pitch and timbre had comparatively little predictive value (between 4 and 4.8 % explained variance). The perception of melody and rhythm requires to track acoustic information over time, whereas pitch and timbre are static auditory stimuli. Thus, the differences in prediction values of the subtests show that rather than static acoustic information, the processing of dynamic cues is of particular importance for emotion perception. The role pitch plays for inferring vocal emotions seems to relate especially to how an individual is able to track its changes over time, i.e. the pitch contour or melody. This mirrors previous findings (Globerson et al., 2013; Greenspon & Montanaro, 2023). Concerning rhythm, it remains unclear how enhanced perception skills transfer to vocal emotion perception. The correlational patterns of the present study provide limited insight as we only manipulated pitch and timbre, with timing held constant across emotions. This may be a potential direction for future research (Swaminathan & Schellenberg, 2020).

The advantages of amateurs and professionals over non-musicians in full and F0-modulated conditions are an example of the musician's benefit in vocal emotion perception, even if there was no main effect of group. With auditory sensitivity rather than music training as the presumable reason for these group differences, the present findings represent a supporting argument for the nature side in the debate whether nature or nurture is responsible for the link between musicality and enhanced vocal emotion perception. A natural predisposition for enhanced auditory sensitivity may make it more likely that a person starts and continues music training (Dmitrieva et al., 2006; Nussbaum & Schweinberger, 2021; Pantev & Herholz, 2011; Schellenberg & Lima, 2024), which would explain why musicians as a group often outperform non-musicians in vocal emotion perception, while also accounting for evidence that musically untrained individuals with high musical perception skills have similar advantages (Correia et al., 2022; Mankel & Bidelman, 2018; M. Martins et al., 2021). Music training undeniably has experience-dependent effects on speech perception at behavioral and neural levels (François et al., 2013; Frey et al., 2019; Moreno et al., 2009; Neves et al., 2022), but such effects do not seem to extend to the perception of vocal emotions.

Interestingly, whereas vocal emotion recognition performance was associated with the objectively measured music perception skills, there were no correlations with the self-rated musical sophistication scores of the Gold-MSI. The missing link with the subscale for music training emphasizes that music training and its amount was not crucial for vocal emotion perception. That there were no associations with self-rated music perception skills either, however, is rather surprising in light of contrasting previous findings where self-rated perceptual (and singing) abilities were positively correlated with emotion recognition accuracy (Correia et al., 2022). It also stands in contrast to the strong link between vocal emotion perception and performance-based measures of music perception skills found in the present study. Questionnaires concerning self-rated perceptual abilities do not necessarily depict an accurate reflection of the actual perceptual skills of participants (Dunning, 2005; Mabe & West, 1982). In that vein, a possible reason for the difference could lie in musicians being more self-critical in regard to their musical skills than would be appropriate for their actual skill level (Hewitt, 2015), especially when compared to non-musicians. The order of tasks in the present study, i.e. the PROMS before the Gold-MSI, may have also played a role. The PROMS is designed to assess a wide range of musical abilities (Zentner & Strauss, 2017); many amateurs reported in debriefing that it was harder than they had initially assumed. To rate their perceptual skills after such a test may have skewed their self-rating. This could be the case especially for amateurs, assuming that professional musicians are more conscious of their own skill level due to constant (self- and external) evaluations and that non-musicians presumably have lower expectations to excel at musical perception skills. In the study by Correia et al. (2022), participants filled out the Gold-MSI before any of the perception tasks, which could be a reason for the contrasting findings.

Concerning preexisting differences, it is further interesting that amateurs reported higher household income than the other two groups. Taking up and continuing music lessons, especially in childhood, is impacted by socioeconomic variables, such as family income and parental support (Corrigall et al., 2013; Schellenberg, 2020b). In the discussed literature, musicians and non-musicians were generally matched in socioeconomic variables or associations remained when controlling for socioeconomic status (e.g., Lima & Castro, 2011; Swaminathan & Schellenberg, 2017), which makes a systematic influence on voice perception unlikely. For our sample, it could be that amateurs had a higher income than non-musicians as a result of socioeconomic differences that also informed their musical activities. In comparison to professional musicians, amateurs may have chosen to pursue careers that are more financially secure than music careers typically tend to be. There is also the fact that a majority of amateurs were full-time university students (due to the fact that most of the choirs and orchestras where recruitment took place were affiliated with local universities) and as such were asked to report the average income of the household they grew up in. Maybe growing up in a household with more income both informed their engagement in musical activities and the choice to pursue a degree with more financial security than a music-related academic degree is associated with.

4.3 Limitations and future directions

The present study is subject to a number of limitations that need to be considered when interpreting the results.

First, the sample is comprised of native or fluent German speakers that were primarily socialized in Western music culture. There is evidence that the language background of listeners influences how they rely on pitch contour and timbre cues to rate affect in emotional voices (Yanushevskaya et al., 2018). Moreover, music style varies as much as culture (Cross, 2008), and can differ greatly in how harmony, pitch or rhythm are utilized (Morrison & Demorest, 2009). Therefore, in order to generalize the present findings to other language backgrounds or musical cultures, further research is needed with more diverse sets of participants.

Second, while we made an effort to make the amateur groups mutually exclusive from each other, it is challenging to find singers who have never played an instrument in any capacity or instrumentalists who have never sung. Instead, participants had to choose between singing vs. playing an instrument when asked about their predominant musical activity. Further, we recruited active members of choirs and orchestras. This did not preclude participants from ever having engaged with another form of music making (e.g., some singers reported to play an instrument). The recruitment in choirs and orchestras also added an element of social interaction to the musical engagement of amateurs that could potentially influence vocal emotion perception. There is some evidence suggesting that making music in a group has socio-emotional benefits (Koelsch, 2013; Schellenberg & Lima, 2024). Moreover, it is conceivable that collective music making may attract more socially inclined individuals. However, we found no differences in personality traits between professionals, amateurs and non-musicians (except for professionals scoring slightly higher on Extraversion than amateurs) and correlational analyses based on personality traits yielded no results which would indicate links to the vocal emotion recognition performance.

Third, we observed slight differences in autistic traits between professional musicians, amateurs and non-musicians. While the overall AQ scores were comparable, professional musicians scored lower than the other two groups on scales of the social communication domain and non-musicians scored lower than both musician groups on the attention to detail domain. This could indicate a link of musicality and autistic traits, which has been found between clinical levels of autism and rare abilities like absolute pitch (Bonnel et al., 2003; Wenhart et al., 2019). However, open questions remain about associations of non-clinical autistic traits with musical expertise (Sivathasan et al., 2022) and how those affect vocal emotion perception.

Fourth, vocal emotion perception was examined using short pseudowords with two syllables, manipulated via parameter-specific voice morphing. Considering the discussed importance of dynamic over static auditory processing, it could be worth exploring whether and how our findings generalize to utterances with longer or shorter lengths, such as (pseudo-)

sentences or single syllables. Furthermore, parameter-specific manipulation via voice morphing is a relatively new tool in voice research. As such, there are open questions about potential side effects and how those could limit the ecological validity of the produced stimuli. For example, a recent study found voice-morphing to reduce the perceived naturalness of voices, but crucially this did not affect the perception of vocal emotions (Nussbaum, Pöhlmann, et al., 2023). As a final limitation regarding the stimuli used in the present study, we focused on the role of pitch and timbre cues. Timing and amplitude may need to be investigated more closely in further research.

Fifth, neither the present study nor the study by Nussbaum, Schirmer and Schweinberger (2023b) included objective measures of cognitive abilities. In order to keep the testing duration to a reasonable length, we only assessed formal education instead. As discussed above, musicality and general cognitive abilities have been linked in the past (Schellenberg, 2001; Schellenberg & Lima, 2024; Vincenzi et al., 2022). Nevertheless, a substantial disadvantage in cognitive abilities of the non-musician group seems unlikely, considering they were largely recruited at university and often either pursued or had completed a PhD.

Sixth and finally, we expected no differences in vocal emotion perception performance between singers and instrumentalists, i.e. predicted null findings. Strictly speaking, Null Hypothesis Significance Testing (NHST) is not appropriate in this case. Rather, one of the most suited methods for statistical analyses of null hypotheses is Bayesian testing (Wei et al., 2022). Given the constraints of a master thesis, we decided to employ NHST nevertheless. For a future publication, the current data may need to be analyzed using a more suited statistical method.

4.4 Conclusion

In conclusion, the present study reinforced the notion that individuals with naturally high musical abilities exhibit benefits regarding vocal emotion perception. While musicians are a very heterogeneous group, neither possible differences between singing vs. playing an instrument nor differences between making music on a professional vs. an amateur level had an empirically observable effect on vocal emotion perception. Furthermore, instead of group differences concerning musical expertise, vocal emotion perception performance was found to be dependent on where individuals were located on a spectrum of naturally predisposed auditory sensitivity and how they utilized pitch contours to infer emotional meaning from speech.

5. References

Anikin, A. (2020). A Moan of Pleasure Should Be Breathy: The Effect of Voice Quality on the Meaning of Human Nonverbal Vocalizations. *Phonetica*, *77*(5), 327–349. https://doi.org/10.1159/000504855

Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., & Armony, J. L. (2015). Fear across the senses: Brain responses to music, vocalizations and facial expressions. *Social Cognitive and Affective Neuroscience*, *10*(3), 399–407. https://doi.org/10.1093/scan/nsu067

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia. A group study of adults afflicted with a music-specific disorder. *Brain*, *125*(2), 238–251. https://doi.org/10.1093/brain/awf028

Baharloo, S., Johnston, P. A., Service, S. K., Gitschier, J., & Freimer, N. B. (1998). Absolute pitch: An approach for identification of genetic and nongenetic components. *American Journal of Human Genetics*, *62*(2), 224–231. https://doi.org/10.1086/301704

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expressions. *Journal of Personality and Social Psychology*, *70*(3), 614–636.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/a:1005653411471

Başkent, D., Fuller, C. D., Galvin, J. J., Schepel, L., Gaudrain, E., & Free, R. H. (2018). Musician effect on perception of spectro-temporally degraded speech, vocal emotion, and music in young adolescents. *The Journal of the Acoustical Society of America*, *143*(5), EL311–EL316. https://doi.org/10.1121/1.5034489

Belin, P., Zatorre, R. J., Lafallie, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–312. https://doi.org/10.1038/35002078

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological*, *57*, 289–300. https://api.semanticscholar.org/CorpusID:45174121

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology*, *2*, 1–12. https://doi.org/10.3389/fpsyg.2011.00094

Bhatara, A., Tirovolas, A. K., Duan, L. M., Levy, B., & Levitin, D. J. (2011). Perception of emotional expression in musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 921–934. https://doi.org/10.1037/a0021922

Bickerton, D. (1990). *Languages and species*. Chicago University Press.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*(1), 100–130. https://doi.org/10.1016/j.cognition.2005.11.007

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion*, *19*(8), 1113–1139. https://doi.org/10.1080/02699930500204250

Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., Deeley, Q., Barker, G. J., Renvall, V., Deoni, S., Gasston, D., Williams, S. C. R., Johnson, M. H., Simmons, A., & Murphy, D. G. M. (2011). Early specialization for voice and emotion processing in the infant brain. *Current Biology*, *21*(14), 1220–1224. https://doi.org/10.1016/j.cub.2011.06.009

Bodner, E., Aharoni, R., & Iancu, I. (2012). The effect of training with music on happiness recognition in social anxiety disorder. *Journal of Psychopathology and Behavioral Assessment*, *34*(4), 458–466. https://doi.org/10.1007/s10862-012-9304-7

Boersma, P. (2018). *Praat: doing phonetics by computer [Computer program]* (6.0.46). http://www.praat.org.

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., & Bonnel, A. M. (2003). Enhanced pitch sensitivity in individuals with autism: A signal detection analysis. *Journal of Cognitive Neuroscience*, *15*(2), 226–235. https://doi.org/10.1162/089892903321208169

Breyer, B., & Bluemke, M. (2016). *Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel)*. https://doi.org/10.6102/zis242

Brown, S. (2000). The “Musilanguage” model of music evolution. In S. Brown, B. Merker, & N. L. Wallin (Eds.), *Theories of Music Origin* (pp. 271–300). MIT Press.

Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, *8*(4), 383–403. https://doi.org/10.1016/j.plrev.2011.10.002

Castro, S. L., & Lima, C. F. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody. *Behavior Research Methods*, *42*(1), 74–81. https://doi.org/10.3758/BRM.42.1.74

Castro, S. L., & Lima, C. F. (2014). Age and Musical Expertise influence emotion recognition in music. *Music Perception*, *32*(2), 125–142. doi:10.1525/mp.2014.32.2.125

Chari, D. A., Barrett, K. C., Patel, A. D., Colgrove, T. R., Jiradejvong, P., Jacobs, L. Y., & Limb, C. J. (2020). Impact of Auditory-Motor Musical Training on Melodic Pattern Recognition in Cochlear Implant Users. *Otology and Neurotology*, *41*(4), e422–e431. https://doi.org/10.1097/MAO.0000000000002525

Chartrand, J. P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters*, *405*(3), 164–167. https://doi.org/10.1016/j.neulet.2006.06.053

Chartrand, J. P., Peretz, I., & Belin, P. (2008). Auditory recognition expertise and domain specificity. *Brain Research*, *1220*, 191–198. https://doi.org/10.1016/j.brainres.2008.01.014

Cheung, Y. L., Zhang, C., & Zhang, Y. (2021). Emotion processing in congenital amusia: the deficits do not generalize to written emotion words. *Clinical Linguistics and Phonetics*, *35*(2), 101–116. https://doi.org/10.1080/02699206.2020.1719209

Christiner, M., & Reiterer, S. M. (2015). A Mozart is not a Pavarotti: singers outperform instrumentalists on foreign accent imitation. *Frontiers in Human Neuroscience*, *9*, 482. https://doi.org/10.3389/fnhum.2015.00482

Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, *352*, 49–69. https://doi.org/10.1016/j.heares.2017.02.006

Correia, A. I., Castro, S. L., Macgregor, C., Müllensiefen, D., Schellenberg, E. G., Correia, A. I., Macgregor, C., Müllensiefen, D., & Schellenberg, E. G. (2022). Enhanced Recognition of Vocal Emotions in Individuals With Naturally Good Musical Abilities. *Emotion*, *22*(5), 894–906. https://doi.org/10.1037/emo0000770

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music Training, Cognition, and Personality. *Frontiers in Psychology*, *4*, 1–10. https://doi.org/10.3389/fpsyg.2013.00222

Coutinho, E., & Dibben, N. (2013). Psychoacoustic cues to emotion in speech prosody and music. *Cognition and Emotion*, *27*(4), 658–684. https://doi.org/10.1080/02699931.2012.732559

Cross, I. (2008). Musicality and the human capacity for culture. *Musicae Scientiae*, *12*(1_suppl), 147–

167. https://doi.org/10.1177/1029864908012001071

Cui, A., & Kuang, J. (2019). The effects of musicality and language background on cue integration in pitch perception. *The Journal of the Acoustical Society of America*, *146*(6), 4086–4096. https://doi.org/10.1121/1.5134442

Curtis, M. E., & Bharucha, J. J. (2010). The Minor Third Communicates Sadness in Speech, Mirroring Its Use in Music. *Emotion*, *10*(3), 335–348. https://doi.org/10.1037/a0017928

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. John Murray.

Darwin, C. (1872). *The expression of the emotions in man and animals*. John Murray.

Dissanayake, E. (2000). Antecedents of the temporal arts in early mother–infant interaction. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 389–410). MIT Press.

Dmitrieva, E. S., Gel'man, V. Y., Zaitseva, K. A., & Orlov, A. M. (2006). Ontogenetic features of the psychophysiological mechanisms of perception of the emotional component of speech in musically gifted children. *Neuroscience and Behavioral Physiology*, *36*(1), 53–62. https://doi.org/10.1007/s11055-005-0162-6

Dunning, D. (2005). *Self-insight: Roadblocks and detours on the path to knowing thyself*. Psychology Press. https://doi.org/10.4324/9780203337998

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased Cortical Representation of the Fingers of the Left Hand in String Players. *Science*, *270*(5234), 305–307. https://doi.org/10.1126/science.270.5234.305

Elmer, S., Dittinger, E., & Besson, M. (2018). One Step Beyond: Musical Expertise and Word Learning. In S. Frühholz & P. Belin (Eds.), *The Oxford Handbook of Voice Perception* (pp. 208–234). Oxford University Press.

Escoffier, N., Zhong, J., Schirmer, A., & Qiu, A. (2013). Emotional expressions in voice and music: Same code, same effect? *Human Brain Mapping*, *34*(8), 1796–1810. https://doi.org/10.1002/hbm.22029

Fisher, R. A., Hoult, A. R., & Tucker, W. S. (2020). A Comparison of Facial Muscle Activation for Vocalists and Instrumentalists. *Journal of Music Teacher Education*, *30*(1), 53–64. https://doi.org/10.1177/1057083720947412

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, *100*(1), 173–215. https://doi.org/10.1016/j.cognition.2005.11.009

Fitch, W. T. (2013). Musical Protolanguage: Darwin's Theory of Language Evolution Revisited. In J. J. Bolhuis & M. Everaert (Eds.), *Birdsong, Speech, and Language: Exploring the Evolution of Mind and Brain* (pp. 489–504). MIT Press. https://doi.org/10.7551/mitpress/9322.003.0032

Francisco, R. M. C. (2021). *Pianos and microphones: Does the type of musical training affect emotion recognition?* [Master's thesis, University of Lisbon]. https://repositorio.ul.pt/bitstream/10451/58438/1/ulfpie058199_tm.pdf

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex*, *23*(9), 2038–2043. https://doi.org/10.1093/cercor/bhs180

Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., Rösler, M., & von Gontard, A. (2007). Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ) - die Kurzversion AQ-k. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, *36*(4), 280–289. https://doi.org/10.1026/1616-3443.36.4.280

Frey, A., François, C., Chobert, J., Velay, J. L., Habib, M., & Besson, M. (2019). Music training positively influences the preattentive perception of voice onset time in children with dyslexia: A longitudinal study. *Brain Sciences*, *9*(4). https://doi.org/10.3390/brainsci9040091

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., & Koelsch, S. (2009). Universal Recognition of Three Basic Emotions in Music. *Current Biology*, *19*(7), 573–576. https://doi.org/10.1016/j.cub.2009.02.058

Frühholz, S., Klaas, H. S., Patel, S., & Grandjean, D. (2015). Talking in fury: The cortico-subcortical network underlying angry vocalizations. *Cerebral Cortex*, *25*(9), 2752–2762. https://doi.org/10.1093/cercor/bhu074

Frühholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions-Towards a unifying neural network perspective of affective sound processing. *Neuroscience and Biobehavioral Reviews*, *68*, 96–110. https://doi.org/10.1016/j.neubiorev.2016.05.002

Fuller, C. D., Galvin, J. J., Maat, B., Başkent, D., & Free, R. H. (2018). Comparison of Two Music Training Approaches on Music and Speech Perception in Cochlear Implant Users. *Trends in Hearing*, *22*, 1–22. https://doi.org/10.1177/2331216518765379

Fuller, C. D., Galvin, J. J., Maat, B., Free, R. H., & Başkent, D. (2014). The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience*, *8*, 1–16. https://doi.org/10.3389/fnins.2014.00179

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*(1), 68–91. https://doi.org/10.1177/0305735696241007

Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to "happy-sad" judgements in equitone melodies. *Cognition and Emotion*, *17*(1), 25–40. https://doi.org/10.1080/02699930302279

Globerson, E., Amir, N., Golan, O., Kishon-Rabin, L., & Lavidor, M. (2013). Psychoacoustic abilities as predictors of vocal emotion recognition. *Attention, Perception, and Psychophysics*, *75*(8), 1799–1810. https://doi.org/10.3758/s13414-013-0518-x

Grandjean, D., Bänziger, T., & Scherer, K. R. (2006). Chapter 12 Intonation as an interface between language and affect. *Progress in Brain Research*, *156*(06), 235–247. https://doi.org/10.1016/S0079-6123(06)56012-1

Greenspon, E. B., & Montanaro, V. (2023). Singing ability is related to vocal emotion recognition: Evidence for shared sensorimotor processing across speech and music. *Attention, Perception, and Psychophysics*, *85*(1), 234–243. https://doi.org/10.3758/s13414-022-02613-0

Grichkovtsova, I., Morel, M., & Lacheret, A. (2012). The role of voice quality and prosodic contour in affective speech perception. *Speech Communication*, *54*(3), 414–429. https://doi.org/10.1016/j.specom.2011.10.005

Habib, M., & Besson, M. (2009). What do music training and musical experience teach us about brain plasticity? *Music Perception*, *26*(3), 279–285. https://doi.org/10.1525/MP.2009.26.3.279

Hallam, S. (2017). The impact of making music on aural perception and language skills: A research synthesis. *London Review of Education*, *15*(3), 388–406. https://doi.org/10.18546/LRE.15.3.05

Halwani, G. F., Loui, P., Rüber, T., & Schlaug, G. (2011). Effects of practice and experience on the arcuate fasciculus: Comparing singers, instrumentalists, and non-musicians. *Frontiers in Psychology*, *2*, 1–9. https://doi.org/10.3389/fpsyg.2011.00156

Herholz, S. C., & Zatorre, R. J. (2012). Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure. *Neuron*, *76*(3), 486–502. https://doi.org/10.1016/j.neuron.2012.10.011

Hewitt, M. P. (2015). Self-Efficacy, Self-Evaluation, and Music Performance of Secondary-Level Band Students. *Journal of Research in Music Education*, *63*(3), 298–313. https://doi.org/10.1177/0022429415595611

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the autism-spectrum quotient (AQ): A study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders*, *38*(8), 1555–1566. https://doi.org/10.1007/s10803-008-0538-x

Holmes, E., Kinghorn, E. E., McGarry, L. M., Busari, E., Griffiths, T. D., & Johnsrude, I. S. (2022). Pitch discrimination is better for synthetic timbre than natural musical instrument timbres despite familiarity. *The Journal of the Acoustical Society of America*, *152*(1), 31–42. https://doi.org/10.1121/10.0011918

Ilie, G., & Thompson, W. F. (2006). A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect. *Music Perception: An Interdisciplinary Journal*, *23*(4), 319–330. https://doi.org/10.1525/9780520912274

International Organization for Standardization. (1975). *Acoustics - Standard tuning frequency (Standard musical pitch)* (ISO Standard No. 16:1975). https://www.iso.org/standard/3601.html

Johnstone, T., & Scherer, K. R. (2000). Vocal Communication of Emotion. In M. Lewis & J. Haviland (Eds.), *The Handbook of Emotion* (pp. 220–235). Guilford.

Juslin, P. N., & Laukka, P. (2003). Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kawahara, H., Morise, M., Banno, H., & Skuk, V. G. (2013). Temporally variable multi-aspect N-way morphing based on interference-free speech representations. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1–10. https://doi.org/10.1109/APSIPA.2013.6694355

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 3933–3936. https://api.semanticscholar.org/CorpusID:14929345

Kawahara, H., & Skuk, V. G. (2018). Voice Morphing. In S. Frühholz & P. Belin (Eds.), *The Oxford Handbook of Voice Perception* (pp. 684–706). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.31

Kim, S. (2015). ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Communications for Statistical Applications and Methods*, *22*(6), 665–674. http://dx.doi.org/10.5351/CSAM.2015.22.6.665

Kirkham, J., Lu, S., Wayland, R., & Kaan, E. (2011). Comparison of Vocalists and Instrumentalists on Lexical Tone Perception and Production Tasks. *17th International Congress of Phonetic Sciences (ICPhS XVII)*, 1098–1101.

Kleber, B., Veit, R., Birbaumer, N., Gruzelier, J., & Lotze, M. (2010). The brain of opera singers: Experience-dependent changes in functional activation. *Cerebral Cortex*, *20*(5), 1144–1152. https://doi.org/10.1093/cercor/bhp177

Koelsch, S. (2013). From Social Contact to Social Cohesion—The 7 Cs. *Music & Medicine*, *5*(4), 204–209. https://doi.org/10.1177/1943862113508588

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. https://doi.org/10.1038/nrn2882

Kraus, N., Hornickel, J., Strait, D. L., Slater, J., & Thompson, E. (2014). Engagement in community music classes sparks neuroplasticity and language development in children from disadvantaged backgrounds. *Frontiers in Psychology*, *5*, 1–9. https://doi.org/10.3389/fpsyg.2014.01403

Kraus, N., & White-Schwoch, T. (2017). Neurobiology of Everyday Communication: What Have We

Learned from Music? *Neuroscientist*, *23*(3), 287–298. https://doi.org/10.1177/1073858416653593

Krishnan, S., Lima, C. F., Evans, S., Chen, S., Guldner, S., Yeff, H., Manly, T., & Scott, S. K. (2018). Beatboxers and Guitarists Engage Sensorimotor Regions Selectively When Listening to the Instruments They can Play. *Cerebral Cortex*, *28*(11), 4063–4079. https://doi.org/10.1093/cercor/bhy208

Lagefoged, P. (1996). *Elements of acoustic phonetics*. University of Chicago Press.

Lappe, C., Trainor, L. J., Herholz, S. C., & Pantev, C. (2011). Cortical plasticity induced by short-term multimodal musical rhythm training. *PLoS ONE*, *6*(6). https://doi.org/10.1371/journal.pone.0021493

Laukka, P., Elfenbein, H. A., Rockstuhl, T., Iraki, F. K., Chui, W., & Althoff, J. (2016). Supplemental Material for The Expression and Recognition of Emotions in the Voice Across Five Nations: A Lens Model Analysis Based on Acoustic Features. *Journal of Personality and Social Psychology*, *111*(5), 686–705. https://doi.org/10.1037/pspi0000066.supp

Law, L. N. C., & Zentner, M. (2012). Assessing Musical Abilities Objectively: Construction and Validation of the Profile of Music Perception Skills. *PLOS ONE*, *7*(12), 1–15. https://doi.org/10.1371/journal.pone.0052508

Lima, C. F., Anikin, A., Monteiro, A. C., Scott, S. K., & Castro, S. L. (2019). Automaticity in the recognition of nonverbal emotional vocalizations. *Emotion*, *19*(2), 219–233. https://doi.org/10.1037/emo0000429

Lima, C. F., Brancatisano, O., Fancourt, A., Müllensiefen, D., Scott, S. K., Warren, J. D., & Stewart, L. (2016). Impaired socio-emotional processing in a developmental music disorder. *Scientific Reports*, *6*, 1–13. https://doi.org/10.1038/srep34911

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, *11*(5), 1021–1031. https://doi.org/10.1037/a0024521

Lolli, S. L., Lewenstein, A. D., Basurto, J., Winnik, S., & Loui, P. (2015). Sound frequency affects speech emotion perception: results from congenital amusia. *Frontiers in Psychology*, *6*, 1–10. https://doi.org/10.3389/fpsyg.2015.01340

Lotze, M., Scheler, G., Tan, H. R. M., Braun, C., & Birbaumer, N. (2003). The musician's brain: Functional imaging of amateurs and professionals during performance and imagery. *NeuroImage*, *20*(3), 1817–1829. https://doi.org/10.1016/j.neuroimage.2003.07.018

Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, *67*(3), 280–296. https://doi.org/10.1037/0021-9010.67.3.280

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, *18*(2), 199–211. https://doi.org/10.1162/jocn.2006.18.2.199

Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(51), 13129–13134. https://doi.org/10.1073/pnas.1811793115

Margulis, E. H., Mlsna, L. M., Uppunda, A. K., Parrish, T. B., & Wong, P. C. M. (2009). Selective neyrophysiologic responses to mysic jn instruinentalists with different listening biographies. *Human Brain Mapping*, *30*(1), 267–275. https://doi.org/10.1002/hbm.20503

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians Detect Pitch Violation in a Foreign Language Better Than Nonmusicians: Behavioral and Electrophysiological Evidence. *Journal of Cognitive Neuroscience*, *19*(9), 1453–1463.

https://doi.org/10.1145/nnnnnnn.nnnnnnn

Martins, I., Lima, C. F., & Pinheiro, A. P. (2022). Enhanced salience of musical sounds in singers and instrumentalists. *Cognitive, Affective and Behavioral Neuroscience*, *22*(5), 1044–1062. https://doi.org/10.3758/s13415-022-01007-x

Martins, M., Pinheiro, A. P., & Lima, C. F. (2021). Does Music Training Improve Emotion Recognition Abilities? A Critical Review. *Emotion Review*, *13*(3), 199–210. https://doi.org/10.1177/17540739211022035

Mill, A., Allik, J., Realo, A., & Valk, R. (2009). Age-Related Differences in Emotion Recognition Ability: A Cross-Sectional Study. *Emotion*, *9*(5), 619–630. https://doi.org/10.1037/a0016562

Mithen, S., Morley, I., Wray, A., Tallerman, M., & Gamble, C. (2006). The Singing Neanderthals: The Origins of Music, Language, Mind and Body. *Cambridge Archaeological Journal*, *16*(1), 97–112. https://doi.org/10.1017/S0959774306000060

Moreno, S., & Bidelman, G. M. (2014). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hearing Research*, *308*, 84–97. https://doi.org/10.1016/j.heares.2013.09.012

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex*, *19*(3), 712–723. https://doi.org/10.1093/cercor/bhn120

Morrison, S. J., & Demorest, S. M. (2009). Cultural constraints on music perception and cognition. *Progress in Brain Research*, *178*, 67–77. https://doi.org/10.1016/S0079-6123(09)17805-6

Mualem, O., & Lavidor, M. (2015). Music education intervention improves vocal emotion recognition. *International Journal of Music Education*, *33*(4), 413–425. https://doi.org/10.1177/0255761415584292

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLOS ONE*, *9*(2), 1–23. https://doi.org/10.1371/journal.pone.0089642

Münzer, S., Berti, S., & Pechmann, T. (2002). Encoding of timbre, speech, and tones: Musicians vs. non-musicians. *Psychologische Beiträge*, *44*, 187–202.

Neves, L., Correia, A. I., Castro, S. L., Martins, D., & Lima, C. F. (2022). Does music training enhance auditory and linguistic processing? A systematic review and meta-analysis of behavioral and brain evidence. *Neuroscience and Biobehavioral Reviews*, *140*, 104777. https://doi.org/10.1016/j.neubiorev.2022.104777

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2008). Hearing of note: An electrophysiologic and psychoacoustic comparison of pitch discrimination between vocal and instrumental musicians. *Psychophysiology*, *45*(6), 994–1007. https://doi.org/10.1111/j.1469-8986.2008.00689.x

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians. *The Journal of the Acoustical Society of America*, *125*(1), 328–338. https://doi.org/10.1121/1.3021309

Nilsonne, Å., & Sundberg, J. (1985). Differences in Ability of Musicians and Nonmusicians to Judge Emotional State from the Fundamental Frequency of Voice Samples. *Music Perception: An Interdisciplinary Journal*, *2*(4), 507–516. https://doi.org/10.2307/40285316

Nolden, S., Rigoulot, S., Jolicoeur, P., & Armony, J. L. (2017). Effects of musical expertise on oscillatory brain activity in response to emotional sounds. *Neuropsychologia*, *103*, 96–105. https://doi.org/10.1016/j.neuropsychologia.2017.07.014

Nussbaum, C., Pöhlmann, M., Kreysa, H., & Schweinberger, S. R. (2023). Perceived Naturalness of Emotional Voice Morphs. *Cognition And*, *37*(4), 721–747.

https://doi.org/10.1080/02699931.2023.2200920

Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2022). Contributions of fundamental frequency and timbre to vocal emotion perception and their electrophysiological correlates. *Social Cognitive and Affective Neuroscience*, *17*(12), 1145–1154. https://doi.org/10.1093/scan/nsac033

Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2023a). Electrophysiological Correlates of Vocal Emotional Processing in Musicians and Non-Musicians. *Brain Sciences*, *13*(11), 1563. https://doi.org/10.3390/brainsci13111563

Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2023b). Musicality – Tuned to the melody of vocal emotions. *British Journal of Psychology*, *00*, 1–20. https://doi.org/10.1111/bjop.12684

Nussbaum, C., & Schweinberger, S. R. (2021). Links Between Musicality and Vocal Emotion Perception. *Emotion Review*, *13*(3), 211–224. https://doi.org/10.1177/17540739211022803

Nussbaum, C., von Eiff, C. I., Skuk, V. G., & Schweinberger, S. R. (2022). Vocal emotion adaptation aftereffects within and across speaker genders: Roles of timbre and fundamental frequency. *Cognition*, *219*, 104967. https://doi.org/10.1016/j.cognition.2021.104967

Palomar-García, M. Á., Zatorre, R. J., Ventura-Campos, N., Bueichekú, E., & Ávila, C. (2017). Modulation of Functional Connectivity in Auditory-Motor Networks in Musicians Compared with Nonmusicians. *Cerebral Cortex*, *27*(5), 2768–2778. https://doi.org/10.1093/cercor/bhw120

Pantev, C., & Herholz, S. C. (2011). Plasticity of the human auditory cortex related to musical training. *Neuroscience and Biobehavioral Reviews*, *35*(10), 2140–2154. https://doi.org/10.1016/j.neubiorev.2011.06.010

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport*, *12*(1), 169–174. https://doi.org/10.1097/00001756-200101220-00041

Paquette, S., Takerkart, S., Saget, S., Peretz, I., & Belin, P. (2018). Cross-classification of musical and vocal emotions in the auditory cortex. *Annals of the New York Academy of Sciences*, *1423*(1), 329–337. https://doi.org/10.1111/nyas.13666

Park, M., Gutyrchik, E., Welker, L., Carl, P., Pöppel, E., Zaytseva, Y., Meindl, T., Blautzik, J., Reiser, M., & Bao, Y. (2015). Sadness is unique: Neural processing of emotions in speech prosody in musicians and non-musicians. *Frontiers in Human Neuroscience*, *8*, 1–8. https://doi.org/10.3389/fnhum.2014.01049

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*, 1–14. https://doi.org/10.3389/fpsyg.2011.00142

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, *308*, 98–108. https://doi.org/10.1016/j.heares.2013.08.011

Patel, A. D., Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception*, *25*(4), 357–368. https://doi.org/10.1525/mp.2008.25.4.357

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing Emotions in a Foreign Language. *Journal of Nonverbal Behavior*, *33*(2), 107–120. https://doi.org/10.1007/s10919-008-0065-7

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology*, *111*, 14–25. https://doi.org/10.1016/j.biopsycho.2015.08.008

Peretz, I. (2008). Musical Disorders. From Behavior To Genes. *Current Directions in Psychological Science*, *17*(5), 329–333. https://doi.org/10.1111/j.1467-8721.2008.00600.x

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, *7*(8), 362–367. https://doi.org/10.1016/S1364-6613(03)00150-5

Piazza, E. A., Theunissen, F. E., Wessel, D., & Whitney, D. (2018). Rapid Adaptation to the Timbre of Natural Sounds. *Scientific Reports*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-32018-9

Pinheiro, A. P., Vasconcelos, M., Dias, M., Arrais, N., & Gonçalves, Ó. F. (2015). The music of language: An ERP investigation of the effects of musical training on emotional prosody processing. *Brain and Language*, *140*, 24–34. https://doi.org/10.1016/j.bandl.2014.10.009

Pralus, A., Fornoni, L., Bouet, R., Gomot, M., Bhatara, A., Tillmann, B., & Caclin, A. (2019). Emotional prosody in congenital amusia: Impaired and spared processes. *Neuropsychologia*, *134*, 107234. https://doi.org/10.1016/j.neuropsychologia.2019.107234

Quinto, L., & Thompson, W. F. (2012). Composing by Listening: A Computer-Assisted System for Creating Emotional Music. *International Journal of Synthetic Emotions*, *3*(2), 48–67. https://doi.org/10.4018/jse.2012070103

R Core Team. (2020). *R: A language and environment for statistical computing*. https://www.r-project.org/

Rammstedt, B., Danner, D., Soto, C. J., & John, O. P. (2020). Validation of the Short and Extra-Short Forms of the Big Five Inventory-2 (BFI-2) and Their German Adaptations. *European Journal of Psychological Assessment*, *36*(1), 149–161. https://doi.org/10.1027/1015-5759/a000481

Rauscher, F. H., & Hinton, S. C. (2011). Music instruction and its diverse extra-musical benefits. *Music Perception*, *29*(2), 215–226. https://doi.org/10.1525/mp.2011.29.2.215

Rigoulot, S., Pell, M. D., & Armony, J. L. (2015). Time course of the influence of musical expertise on the processing of vocal and musical sounds. *Neuroscience*, *290*, 175–184. https://doi.org/10.1016/j.neuroscience.2015.01.033

Rø, M. H., Behne, D. M., & Wang, Y. (2006). The effects of musical experience on linguistic pitch perception: A comparison of Norwegian professional singers and instrumentalists. *The Journal of the Acoustical Society of America*, *120*(5), 3168–3168. https://doi.org/10.1121/1.4787906.

Rogenmoser, L., Kernbach, J., Schlaug, G., & Gaser, C. (2018). Keeping brains young with making music. *Brain Structure and Function*, *223*(1), 297–305. https://doi.org/10.1007/s00429-017-1491-2

Sachs, M. E., Habibi, A., Damasio, A., & Kaplan, J. T. (2018). Decoding the neural signatures of emotions expressed through sound. *NeuroImage*, *174*, 1–10. https://doi.org/10.1016/j.neuroimage.2018.02.058

Sala, G., & Gobet, F. (2020). Cognitive and academic benefits of music training with children : A multilevel meta-analysis. *Memory & Cognition*, *48*(8), 1429–1441. https://doi.org/10.3758/s13421-020-01060-2

Sares, A. G., Foster, N. E. V., Allen, K., & Hyde, K. L. (2018). Pitch and time processing in speech and tones: The effects of musical training and attention. *Journal of Speech, Language, and Hearing Research*, *61*(3), 496–509. https://doi.org/10.1044/2017_JSLHR-S-17-0207

Scheiner, E., & Fischer, J. (2011). Emotion Expression: The Evolutionary Heritage in the Human Voice. In W. Welsch, W. J. Singer, & A. Wunder (Eds.), *Interdisciplinary Anthropology: Continuing Evolution of Man* (pp. 105–129). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-11668-1_5

Schellenberg, E. G. (2001). Music and Nonmusical Abilities. *Annals of the New York Academy of Sciences*, *930*(1), 355–371. https://doi.org/10.1111/j.1749-6632.2001.tb05744.x

Schellenberg, E. G. (2016). Music training and nonmusical abilities. *The Oxford Handbook of Music*

Psychology, *2*, 415–429.

Schellenberg, E. G. (2020a). Correlation = causation? Music training, psychology, and neuroscience. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(4), 475–480. https://doi.org/10.1037/aca0000263

Schellenberg, E. G. (2020b). Music training, individual differences, and plasticity. In M. S. C. Thomas, D. Mareschal, & I. Dumontheil (Eds.), *Educational neuroscience: Development across the life span* (1st ed., pp. 413–438). Routledge.

Schellenberg, E. G., & Lima, C. F. (2024). Music Training and Nonmusical Abilities. *Annual Review of Psychology*, *75*, 87–128. https://doi.org/10.1146/annurev-psych-032323-051354

Scherer, K. R. (1986). Vocal Affect Expression: A Review and a Model for Future Research. *Psychological Bulletin*, *99*(2), 143–165.

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, *9*(3), 235–248. https://doi.org/10.1016/S0892-1997(05)80231-0

Scherer, K. R. (2018). Acoustic Patterning of Emotion Vocalizations. In S. Frühholz & P. Belin (Eds.), *The Oxford Handbook of Voice Perception* (pp. 60–92). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.4

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, *32*(1), 76–92. https://doi.org/10.1177/0022022101032001009

Scherer, K. R., Trznadel, S., Fantini, B., & Sundberg, J. (2017). Recognizing Emotions in the Singing Voice. *Psychomusicology: Music, Mind, and Brain*, *27*, 244–255. https://doi.org/10.1037/pmu0000193

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24–30. https://doi.org/10.1016/j.tics.2005.11.009

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2005). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research*, *24*(3), 442–452. https://doi.org/10.1016/j.cogbrainres.2005.02.022

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341–349. https://doi.org/10.1111/1469-8986.00172.x

Schutz, M. (2017). Acoustic constraints and musical consequences: Exploring composers' use of cues for musical emotion. *Frontiers in Psychology*, *8*, 1–10. https://doi.org/10.3389/fpsyg.2017.01402

Shahin, A. J., Roberts, L. E., Chau, W., Trainor, L. J., & Miller, L. M. (2008). Music training leads to the development of timbre-specific gamma band activity. *NeuroImage*, *41*(1), 113–122. https://doi.org/10.1016/j.neuroimage.2008.01.067

Sharp, A., Houde, M. S., Bacon, B. A., & Champoux, F. (2019). Musicians show better auditory and tactile identification of emotions in music. *Frontiers in Psychology*, *10*, 1–7. https://doi.org/10.3389/fpsyg.2019.01976

Shorey, A. E., King, C. J., Whiteford, K. L., & Stilp, C. E. (2024). Musical training is not associated with spectral context effects in instrument sound categorization. *Attention, Perception, and Psychophysics*, *86*, 991–1007. https://doi.org/10.3758/s13414-023-02839-6

Sivathasan, S., Philibert-Lignières, G., & Quintin, E. M. (2022). Individual Differences in Autism Traits, Personality, and Emotional Responsiveness to Music in the General Population. *Musicae Scientiae*, *26*(3), 538–557. https://doi.org/10.1177/1029864920988160

Stephey, L. A., & Moore, T. R. (2008). Experimental investigation of an American five-string banjo. *The Journal of the Acoustical Society of America*, *124*(5), 3276–3283. https://doi.org/10.1121/1.2982371

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Strait, D. L., & Kraus, N. (2011). Playing music for the smarter ear: cognitive, perceptual and neurobiological evidence. *Music Perception*, *29*(2), 133–147. https://doi.org/10.1525/mp.2011.29.2.133

Strait, D. L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience and neural efficiency - Effects of training on subcortical processing of vocal expressions of emotion. *European Journal of Neuroscience*, *29*(3), 661–668. https://doi.org/10.1111/j.1460-9568.2009.06617.x

Swaminathan, S., & Schellenberg, E. G. (2017). Musical competence and phoneme perception in a foreign language. *Psychonomic Bulletin and Review*, *24*(6), 1929–1934. https://doi.org/10.3758/s13423-017-1244-5

Swaminathan, S., & Schellenberg, E. G. (2020). Musical ability, music training, and language ability in childhood. *Journal of Experimental Psychology: Learning Memory and Cognition*, *46*(12), 2340–2348. https://doi.org/10.1037/xlm0000798

Symons, A. E., & Tierney, A. T. (2024). Supplemental Material for Musical Experience Is Linked to Enhanced Dimension-Selective Attention to Pitch and Increased Primary Weighting During Suprasegmental Categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(2), 189–203. https://doi.org/10.1037/xlm0001217.supp

Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PLoS ONE*, *12*(10), 1–21. https://doi.org/10.1371/journal.pone.0186773

Tervaniemi, M. (2009). Musicians - Same or different. *Annals of the New York Academy of Sciences*, *1169*, 151–156. https://doi.org/10.1111/j.1749-6632.2009.04591.x

Thompson, W. F. (2009). *Music, thought and feeling: Understanding the psychology of music*. Oxford University Press. http://pom.sagepub.com/cgi/doi/10.1177/0305735609339477

Thompson, W. F., & Balkwill, L.-L. (2010). Cross-cultural similarities and differences. In P. N. Juslin & J. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 755–788). Oxford University Press. https://doi.org/10.1093/acprof

Thompson, W. F., & Balkwill, L. L. (2006). Decoding speech prosody in five languages. *Semiotica*, *158*(1/4), 407–424. https://doi.org/10.1515/SEM.2006.017

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(46), 19027–19032. https://doi.org/10.1073/pnas.1210344109

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding Speech Prosody: Do Music Lessons Help? *Emotion*, *4*(1), 46–64. https://doi.org/10.1037/1528-3542.4.1.46

Toh, X. R., Tan, S. H., Wong, G., Lau, F., & Wong, F. C. K. (2023). Enduring musician advantage among former musicians in prosodic pitch perception. *Scientific Reports*, *13*(1), 2657. https://doi.org/10.1038/s41598-023-29733-3

Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience*, *6*(7), 669–673.

https://doi.org/10.1038/nn1084

Trimmer, C. G., & Cuddy, L. L. (2008). Emotional Intelligence, Not Music Training, Predicts Recognition of Emotional Speech Prosody. *Emotion*, *8*(6), 838–849. https://doi.org/10.1037/a0014080

Vincenzi, M., Correia, A. I., Vanzella, P., Pinheiro, A. P., Lima, C. F., & Schellenberg, E. G. (2022). Associations Between Music Training and Cognitive Abilities: The Special Case of Professional Musicians. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000481

Von Eiff, C. I., Skuk, V. G., Zäske, R., Nussbaum, C., Frühholz, S., Feuer, U., Guntinas-Lichius, O., & Schweinberger, S. R. (2022). Parameter-Specific Morphing Reveals Contributions of Timbre to the Perception of Vocal Emotions in Cochlear Implant Users. *Ear and Hearing*, *43*(4), 1178–1188. https://doi.org/10.1097/AUD.0000000000001181

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wei, Z., Yang, A., Rocha, L., Miranda, M. F., & Nathoo, F. S. (2022). A Review of Bayesian Hypothesis Testing and Its Practical Implementations. *Entropy*, *24*(2). https://doi.org/10.3390/e24020161

Weijkamp, J., & Sadakata, M. (2017). Attention to affective audio-visual information: Comparison between musicians and non-musicians. *Psychology of Music*, *45*(2), 204–215. https://doi.org/10.1177/0305735616654216

Weiss, M. W., Schellenberg, G. E., Trehub, S. E., & Dawber, E. J. (2015). Enhanced processing of vocal melodies in childhood. *Developmental Psychology*, *51*(3), 370–377. https://doi.org/10.1037/a0038784

Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the Way She Sings: Enhanced Memory for Vocal Melodies. *Psychological Science*, *23*(10), 1074–1078. https://doi.org/10.1177/0956797612442552

Weiss, M. W., Vanzella, P., Schellenberg, E. G., & Trehub, S. E. (2015). Pianists exhibit enhanced memory for vocal melodies but not piano melodies. *Quarterly Journal of Experimental Psychology*, *68*(5), 866–877. https://doi.org/10.1080/17470218.2015.1020818

Wenhart, T., Bethlehem, R. A. I., Baron-Cohen, S., & Altenmüller, E. (2019). Autistic traits, resting-state connectivity, and absolute pitch in professional musicians: Shared and distinct neural features. *Molecular Autism*, *10*, 1–18. https://doi.org/10.1186/s13229-019-0272-6

Yanushevskaya, I., Gobl, C., & Ní Chasaide, A. (2018). Cross-language differences in how voice quality and f contours map to affect . *The Journal of the Acoustical Society of America*, *144*(5), 2730–2750. https://doi.org/10.1121/1.5066448

Zatorre, R. J. (2003). Absolute pitch: a model for understanding the influence of genes and development on neural and cognitive function. *Nature Neuroscience*, *6*(7), 692–695. https://doi.org/10.1038/nn1085

Zentner, M., & Strauss, H. (2017). Assessing musical ability quickly and objectively: development and validation of the Short-PROMS and the Mini-PROMS. *Annals of the New York Academy of Sciences*, *1400*(1), 33–45. https://doi.org/https://doi.org/10.1111/nyas.13410

Zhang, J. D., Susino, M., McPherson, G. E., & Schubert, E. (2020). The definition of a musician in music psychology: A literature review and the six-year rule. *Psychology of Music*, *48*(3), 389–409. https://doi.org/10.1177/0305735618804038

Zhang, Y., Geng, T., & Zhang, J. (2018). Emotional prosody perception in Mandarin-speaking congenital amusics. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2018-Septe*(September), 2196–2200. https://doi.org/10.21437/Interspeech.2018-91

A. Supplemental Tables

Table A.1

List of reported instruments by amateurs (singers and instrumentalists)

Singers		**Instrumentalists**	
Gesang *(singing)*	37	Violine *(violin)*	10
Klavier *(piano)*	3	Posaune *(trombone)*	6
Gitarre *(guitar)*	2	Klarinette *(clarinet)*	4
Violine *(violin)*	1	Cello *(cello)*	4
Trompete *(trumpet)*	1	Trompete *(trumpet)*	4
Klavier, Cello *(piano, cello)*	1	Waldhorn (*horn*)	2
		Bratsche (*viola*)	2
		Schlagzeug / Schlagwerk (*drums*)	2
		E-Bass (*bass guitar*)	1
		Gitarre (*guitar*)	1
		Klavier (*piano*)	1
		Querflöte (*flute*)	1
		Saxophon (*saxophone*)	1
		Tuba (*tuba*)	1
		Violoncello (*cello*)	1
		Fagott (*bassoon*)	1
		Bariton (*horn*)	1

Table A.2

List of reported instruments by professionals and non-musicians

Professionals		**Non-musicians**	
Klavier (*piano*)	11	Gesang (*singing*)	5
Orgel (*organ*)	7	Klavier (*piano*)	3
Gesang (*singing*)	7	Violine (*violin*)	2
Gitarre (*guitar*)	3	Gitarre (*guitar*)	1
Violine (*violin*)	2	Blockflöte (*flute*)	1
Klarinette (*clarinet*)	2	Tamburin (*tambourine*)	1
Violoncello (*cello*)	1	Fürst-Pless-Horn (*horn*)	1
Kontrabass (*double bass*)	1		
Blockflöte (*flute*)	1		
Oboe (*oboe*)	1		
Querflöte (*flute*)	1		
Trompete (*trumpet*)	1		
Dudelsack (*bagpipe*)	1		

Note. This data were taken from Nussbaum, Schirmer & Schweinberger (2023b), Supplemental Tables (see https://doi.org/10.17605/OSF.IO/3JKCQ).

Table A.3

Socioeconomic background of amateurs (singers and instrumentalists)

Income (€)			**Education**			**Degree**		
	S	**I**		**S**	**I**		**S**	**I**
< 1750	5	4	Keine (*none*)	0	0	Keine (*none*)	0	1
1750-2500	6	7	Schüler (*pupil*)	0	1	Schüler (*pupil*)	0	0
2500-3500	11	10	Hauptschule (*secondary school)*	0	0	inAusbildung (*under training*)	15	17
3500-5000	17	9	Mittelschule (*secondary school*)	0	0	Lehre (*traineeship*)	0	3
> 5000	6	13	Fachschule (*technical college*)	1	1	Fachschule (*technical coll*ege)	1	0
			Abitur (*A-levels*)	44	41	Meister (*master as craftsmen*)	0	0
						Bachelor (*bachelor*)	10	4
						Fachhochschule (*polytechnic degree*)	1	2
						Master/Diplom (*Master/Diploma*)	14	10
						Promotion (*PhD*)	4	6
$\chi^2(4) = 5.23$, p = .264			$\chi^2(2) = 1.06$, p = .588			$\chi^2(7) = 9.06$, p = .249		

Note. This table presents the number of individuals belonging to different income, education, and degree categories. We tested group differences between singers (S) and instrumentalists (I) using a Chi-square test and show the results in the last line of this table. Please note that the response options “Education” (i.e. the type of school) and “Degree” (i.e. the highest professional qualification) were tailored to the German educational system and are therefore difficult to translate. Further, please note that “Fachschule” and “Abitur” are similar as they both enable a person to pursue a university degree (with a few more constrains for a “Fachschule” degree). We therefore consider the trend observed for the “Education” factor merely as an artefact of the response format. S = Singers, I = Instrumentalists.

Table A.4

Socioeconomic background of participants (professionals, amateurs and non-musicians)

Income (€)				**Education**				**Degree**			
	P	**A**	**C**		**P**	**A**	**C**		**P**	**A**	**C**
< 1750	13	9	7	Keine (*none*)	0	1	0	Keine (*none*)	1	1	0
1750-2500	7	13	8	Schüler (*pupil*)	0	0	0	Schüler (*pupil*)	0	0	0
2500-3500	9	21	15	Hauptschule (*secondary school)*	0	0	0	inAusbildung (*under training*)	6	32	6
3500-5000	5	26	6	Mittelschule (*secondary school*)	1	0	0	Lehre (*traineeship*)	2	3	1
> 5000	6	19	2	Fachschule (*technical college*)	0	2	4	Fachschule (*technical coll*ege)	1	1	4
				Abitur (*A-levels*)	39	85	34	Meister (*master as craftsmen*)	0	0	1
								Bachelor (*bachelor*)	8	14	4
								Fachhochschule (*polytechnic degree*)	4	3	2
								Master/Diplom (*Master/Diploma*)	15	24	16
								Promotion (*PhD*)	3	10	4
$\chi^2(8) = 20.19$, $p = .01$				$\chi^2(6) = 11.11$, $p = .085$				$\chi^2(16) = 24.04$, $p = .089$			

Note. This table presents the number of individuals belonging to different income, education, and degree categories. We tested group differences between professionals (P), amateurs (A) and non-musicians (C) using a Chi-square test and show the results in the last line of this table. Please note that the response options “Education” (i.e. the type of school) and “Degree” (i.e. the highest professional qualification) were tailored to the German educational system and are therefore difficult to translate. Further, please note that “Fachschule” and “Abitur” are similar as they both enable a person to pursue a university degree (with a few more constrains for a “Fachschule” degree). We therefore consider the trend observed for the “Education” factor merely as an artefact of the response format. P = Professionals, A = Amateurs, C = Controls/Non-Musicians.

Table A.5

Summary of response key mappings to emotions

	"d"	**"f"**	**"j"**	**"k"**
CB 1	happiness	pleasure	sadness	fear
CB 2	sadness	fear	happiness	pleasure
CB 3	pleasure	happiness	fear	sadness
CB 4	fear	sadness	pleasure	happiness

Note. Participants were instructed explicitly to press the keys "d" and "f" with their left index- and middle-finger and the keys "j" and "k" with their right index- and middle-finger. CB = counterbalancing condition. This data were taken from Nussbaum, Schirmer & Schweinberger (2023b), Supplemental Tables (see https://doi.org/10.17605/OSF.IO/3JKCQ).

Table A.6

Participant assignment to different response key mappings (singers and instrumentalists)

	Singers	**Instrumentalists**
CB 1	11	10
CB 2	10	12
CB 3	16	8
CB 4	8	13

Note. CB = counterbalancing condition.

Table A.7

Participant assignment to different response key mappings (professionals, amateurs and non-musicians)

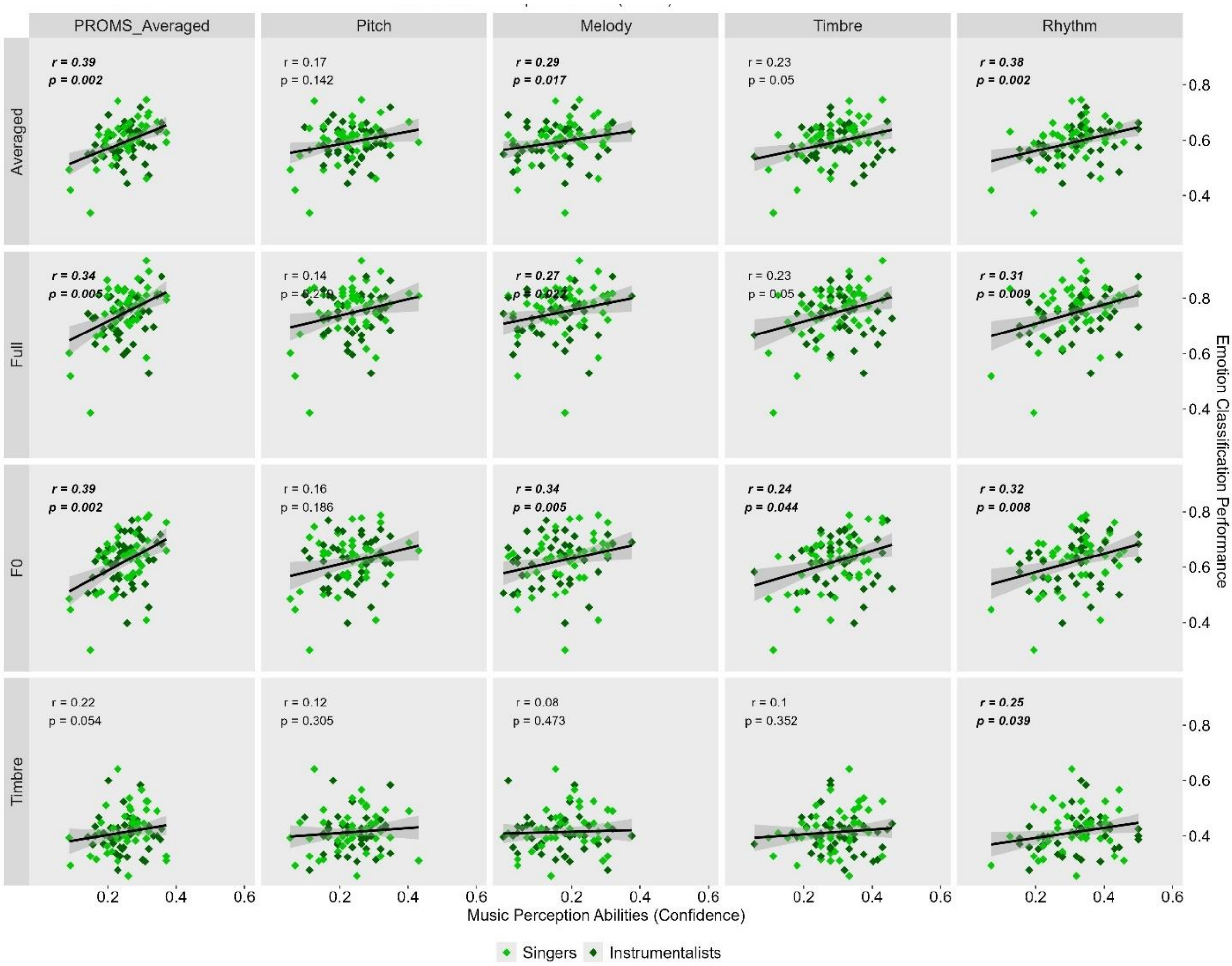
	Professionals	**Amateurs**	**Non-musicians**
CB 1	10	21	7
CB 2	11	22	7
CB 3	8	24	15
CB 4	14	21	9

Note. CB = counterbalancing condition. Data of professionals and non-musicians were taken from Nussbaum, Schirmer & Schweinberger (2023b), Supplemental Tables (see https://doi.org/10.17605/OSF.IO/3JKCQ).

B. Supplemental Figures

Figure B.1

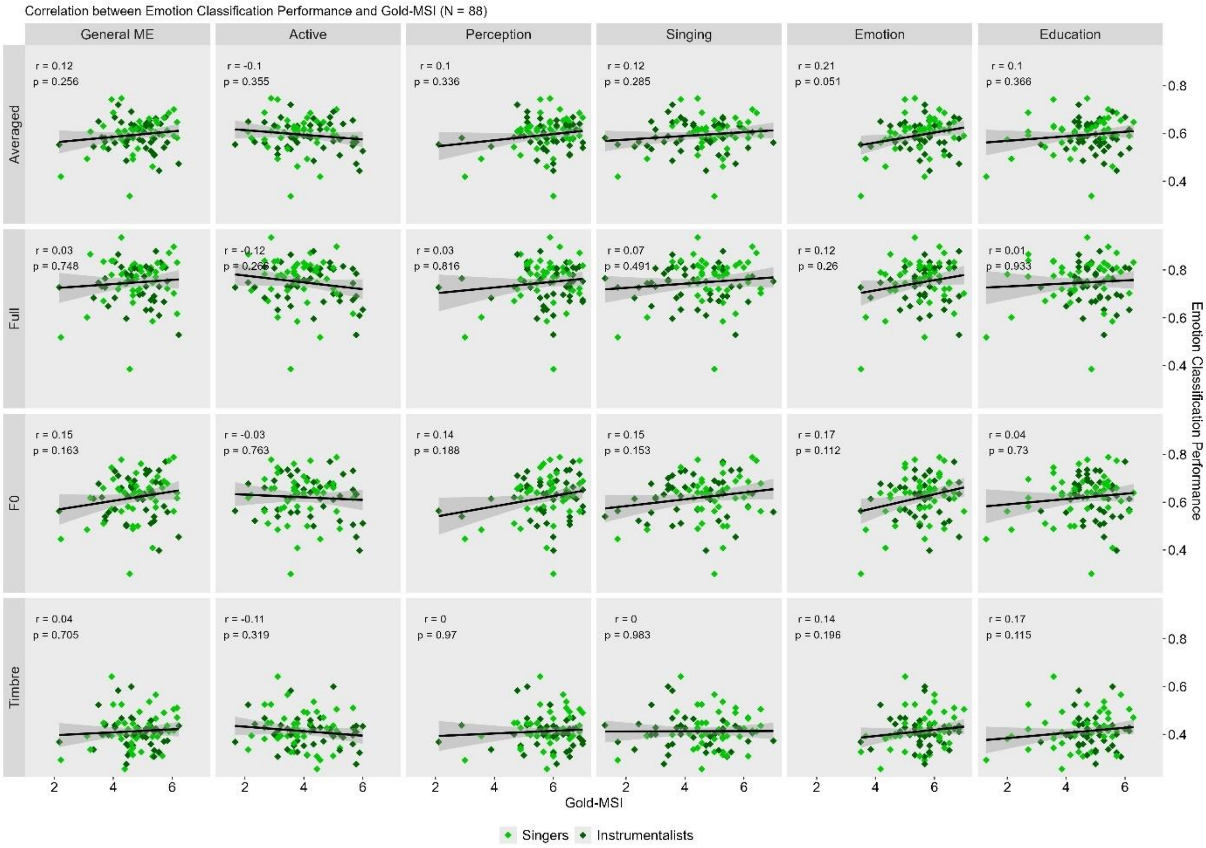
Correlations between vocal emotion recognition and music perception performance (PROMS) for singers and instrumentalists



Note. The x-axis shows the different subtests of the PROMS (Pitch, Melody, Timbre, and Rhythm) as well as the averaged performance across all subtests (PROMS_Averaged). The y-axis shows the vocal emotion classification performance separately for each Morph Type (Full, F0 and Timbre) and averaged across Morph Types (Averaged). p-values were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995; false discovery rate set to 0.05, total number of tests = 20).

Figure B.2

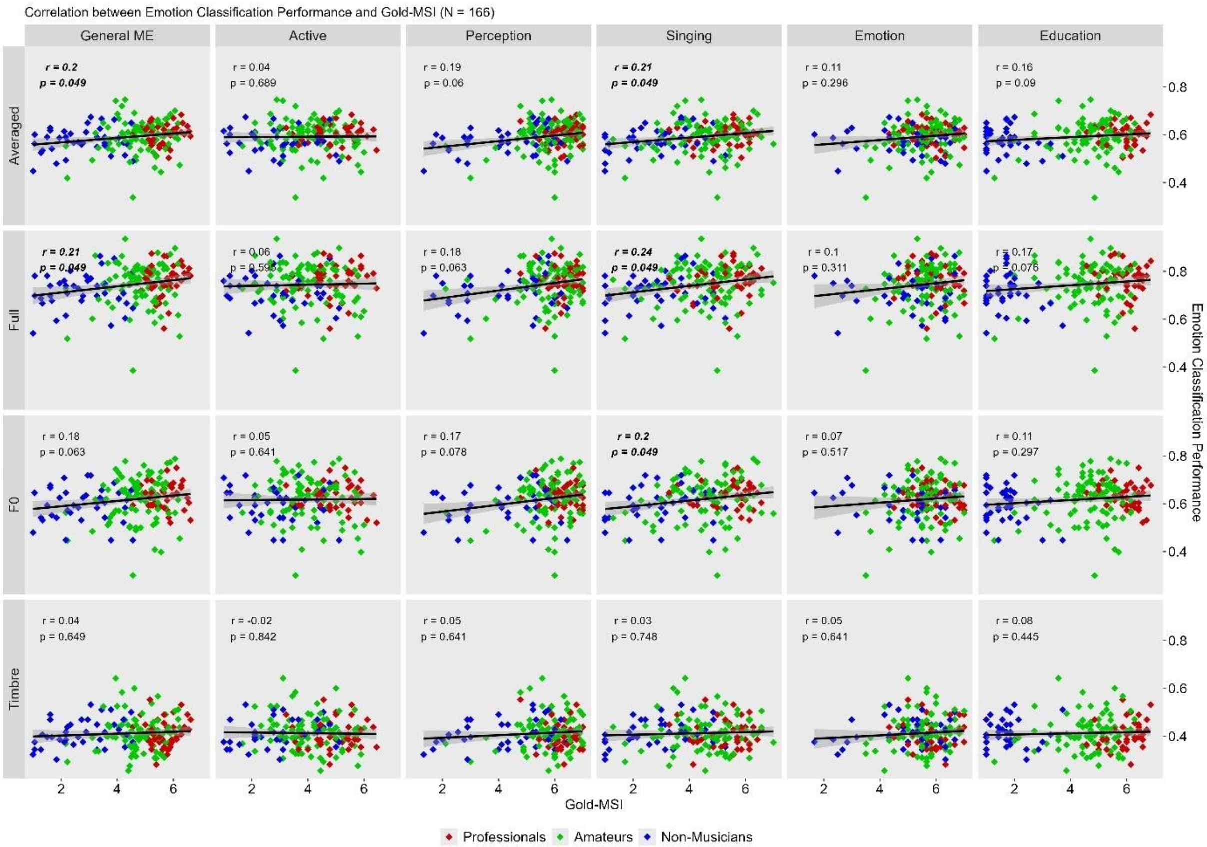
Correlations between vocal emotion recognition and self-rated music skills (Gold-MSI) for singers and instrumentalists



Note. The x-axis shows the different subscores of the Gold-MSI. The y-axis shows the vocal emotion classification performance separately for each Morph Type and averaged across Morph Types (Averaged). p-values were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995; false discovery rate set to 0.05, number of tests = 24).

Figure B.3

Correlations between vocal emotion recognition and self-rated music skills (Gold-MSI) for professionals, amateurs and non-musicians



Note. The x-axis shows the different subscores of the Gold-MSI. The y-axis shows the vocal emotion classification performance separately for each Morph Type and averaged across Morph Types (Averaged). p-values were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995; false discovery rate set to 0.05, number of tests = 24).

C. Supplemental Analyses

Table C.1

Post-hoc tests on the AQ for professionals vs. non-musicians

	Pro-fessionals	**Non-musicians**					
	M (SD)	*M (SD)*	*t*	*df*[a]	*p*	*Cohens d*	
AQ							
Total	15.7 (4.98)	17.58 (6.41)	-1.44	69.83	.154	-0.34 [-0.82, 0-13]	
Attention to Detail	**5.43 (2.04)**	**4.32 (2.01)**	2.42	75.87	.018	0.56 [0.09, 1.01]	*
Social	**10.28 (4.70)**	**13.26 (6.51)**	-2.32	67.08	.024	-0.57 [-1.05, -0.08]	*
Social Skills	**1.48 (1.68)**	**2.61 (2.63)**	-2.25	62.40	.028	-0.57 [-1.07, -0.06]	*
Communication	1.85 (1.61)	2.39 (1.73)	-1.44	74.83	.155	-0.33 [-0.79, 0.13]	
Imagination	2.18 (1.52)	2.87 (1.95)	-1.75	69.92	.085	-0.42 [-0.89, 0.06]	
Attention Switching	4.78 (1.91)	5.39 (1.92)	-1.43	75.75	.158	-0.33 [-0.78, 0.13]	

[a] Note that original degrees of freedom were 76 but were corrected due to unequal variance.

Table C.2

Post-hoc tests on the AQ for professionals vs. amateurs

	Pro-fessionals	**Amateurs**					
	M (SD)	*M (SD)*	*t*	*df*[a]	*p*	*Cohens d*	
AQ							
Total	**15.7 (4.98)**	**18.73 (7.40)**	-2.72	107.7	.008	-0.52 [-0.91, -0.14]	**
Attention to Detail	5.43 (2.04)	5.51 (2.42)	-0.21	88.61	.835	-0.04 [-0.46, 0.37]	
Social	**10.28 (4.70)**	**13.22 (6.49)**	-2.90	101.8	.005	-0.57 [-0.97, -0.18]	**
Social Skills	**1.48 (1.68)**	**2.74 (2.49)**	-3.36	107.8	.001	-0.65 [-1.03, -0.26]	**
Communication	1.85 (1.61)	2.49 (2.12)	-1.88	97.42	.063	-0.38 [-0.78, 0.02]	
Imagination	2.18 (1.52)	2.66 (1.8)	-1.57	89.18	.120	-0.33 [-0.75, 0.09]	
Attention Switching	4.78 (1.91)	5.33 (2.06)	-1.48	80.83	.142	-0.33 [-0.77, 0.11]	

[a] Note that original degrees of freedom were 126 but were corrected due to unequal variance.

Table C.3

Post-hoc tests on the AQ for amateurs vs. non-musicians

	Amateurs	**Non-musicians**					
	M (SD)	*M (SD)*	*t*	*df*[a]	*p*	*Cohens d*	
AQ							
Total	18.73 (7.40)	17.58 (6.41)	-0.88	80.44	.382	-0.2 [-0.63, 0.24]	
Attention to Detail	**5.51 (2.42)**	**4.32 (2.01)**	-2.87	83.53	.005	-0.63 [-1.07, -0.19]	**
Social	13.22 (6.49)	13.26 (6.51)	0.04	70.10	.970	0.01 [-0.46, 0.48]	
Social Skills	2.74 (2.49)	2.61 (2.63)	-0.27	67.11	.791	-0.06 [-0.54, 0.41]	
Communication	2.49 (2.12)	2.39 (1.73)	-0.26	85.06	.795	-0.33 [-0.75, 0.09]	
Imagination	2.66 (1.8)	2.87 (1.95)	0.57	65.93	.574	0.14 [-0.34, 0.62]	
Attention Switching	5.33 (2.06)	5.39 (1.92)	0.17	74.89	.865	0.04 [-0.41, 0.49]	

[a] Note that original degrees of freedom were 124 but were corrected due to unequal variance.

Table C.4

Post-hoc tests on the Gold-MSI for professionals vs. non-musicians

	Pro-fessionals	**Non-musicians**					
	M (SD)	*M (SD)*	*t*	*df*[a]	*P*	*Cohens d*	
Gold-MSI							
General ME	**5.68 (0.50)**	**2.74 (1.07)**	15.45	51.63	<.001	4.30 [3.30, 5.28]	***
Active Engagement	**4.94 (0.81)**	**2.95 (1.19)**	8.55	64.53	<.001	2.13 [1.51, 2.73]	***
Formal Education	**5.95 (0.56)**	**1.71 (0.68)**	30.10	71.67	<.001	7.11 [5.85, 8.36]	***
Emotion	**5.88 (0.73)**	**4.95 (1.32)**	3.79	56.87	<.001	1.00 [0.45, 1.55]	***
Singing	**5.34 (0.83)**	**2.84 (1.26)**	10.3	63.49	<.001	2.59 [1.91, 3.25]	***
Perception	**6.31 (0.51)**	**4.22 (1.49)**	8.19	45.10	<.001	2.44 [1.66, 3.20]	***

[a] Note that original degrees of freedom were 76 but were corrected due to unequal variance.

Table C.5

Post-hoc tests on the Gold-MSI for professionals vs. amateurs

	Pro-fessionals	**Amateurs**					
	M (SD)	*M (SD)*	*t*	*df*[a]	*P*	*Cohens d*	
Gold-MSI							
General ME	**5.68 (0.50)**	**4.76 (0.82)**	7.80	116.1	<.001	1.45 [1.04, 1.85]	***
Active Engagement	**4.94 (0.81)**	**4.02 (1.00)**	5.54	91.98	<.001	1.16 [0.71, 1.59]	***
Formal Education	**5.95 (0.56)**	**4.66 (0.96)**	9.54	118.5	<.001	1.75 [1.33, 2.17]	***
Emotion	**5.88 (0.73)**	**5.55 (0.78)**	2.29	80.76	.025	0.51 [0.06, 0.95]	*
Singing	**5.34 (0.83)**	**4.59 (1.19)**	4.08	105	<.001	0.80 [0.40, 1.19]	***
Perception	**6.31 (0.51)**	**5.75 (0.92)**	4.42	121.3	<.001	0.80 [0.43, 1.17]	***

[a] Note that original degrees of freedom were 126 but were corrected due to unequal variance.

Table C.6

Post-hoc tests on the Gold-MSI for amateurs vs. non-musicians

	Amateurs	Non-musicians					
	M (SD)	*M (SD)*	*t*	*df*[a]	*p*	*Cohens d*	
Gold-MSI							
General ME	**4.76 (0.82)**	**2.74 (1.07)**	-10.41	56.77	<.001	-2.76 [-3.48, -2.03]	***
Active Engagement	**4.02 (1.00)**	**2.95 (1.19)**	-4.81	60.26	<.001	-1.24 [-1.79, -0.68]	***
Formal Education	**4.66 (0.96)**	**1.71 (0.68)**	-19.67	97.53	<.001	-3.98 [-4.66, -3.29]	***
Emotion	**5.55 (0.78)**	**4.95 (1.32)**	-2.59	48.55	.013	-0.74 [-1.32, -0.16]	*
Singing	**4.59 (1.19)**	**2.84 (1.26)**	-7.30	66.78	<.001	-1.79 [-2.35, -1.21]	***
Perception	**5.75 (0.92)**	**4.22 (1.49)**	-5.85	49.64	<.001	-1.66 [-2.30, -1.01]	***

[a] Note that original degrees of freedom were 124 but were corrected due to unequal variance.

C.7: Analysis of Variance for PROMS (professionals, amateurs, and non-musicians)

In the **PROMS**, an ANOVA with PROMS-subtest (Pitch, Melody, Timbre, Rhythm) as repeated measures factor and Group (professionals, amateurs, non-musicians) as between subject factor revealed significant main effects of **Group** (F(2, 163) = 31.33, p < .001, ω^2 = .27, 95%-CI [0.16, 0.37]) and **Subtest** (F(3, 489) = 136.22, p < .001, ω^2 = .45, 95%-CI [0.39, 0.50]) as well as a significant interaction of **Group x Subtest** (F(6, 489) = 4.54, p < .001, ω^2 = .04, 95%-CI [0.01, 0.07]).

Post-hoc tests on the Group main effect showed that professionals (M = 0.29 ± 0.01 SEM) outperformed amateurs (M = 0.25 ± 0.01; |t(375.38)| = 4.09, p < .001, d = 0.42 [0.22, 0.63]) and non-musicians (M = 0.19 ± 0.01; |t(281.45)| = 8.25, p < .001, d = 0.98 [0.74, 1.23]). Amateurs in turn outperformed non-musicians (|t(271.26)| = 5.38, p < .001, d = 0.65 [0.41, 0.90]).

The main effect of Group was found for all subtests separately (all Fs(2, 163) ≥ 4.89, ps ≤ .009; for details see Table 2), with professionals outperforming non-musicians in all subtests (all |ts| ≥ 2.99, ps ≤ .004). Professionals also performed better than amateurs in the Pitch and Melody subtest (Pitch: |t(97.32)| = 2.57, p = .012, d = 0.55 [0.12, 0.98]; Melody: |t(95.24)| = 4.42, p < .001, d = 0.91 [0.48, 1.33]), whereas there were no differences in the Timbre and Rhythm subtests (ps ≥ .09). Amateurs performed better than non-musicians in the Pitch, Melody and Rhythm subtest (Pitch: |t(81.21)| = 4.39, p < .001, d = 0.97 [0.51, 1.43]; Melody: |t(91.34)| = 5.65, p < .001, d = 1.18 [0.74, 1.62]; Rhythm: |t(80.84)| = 3.16, p = .002, d = 0.7 [0.25, 1.15]), but not in the Timbre subtest (p = .064; for all post-hoc tests see Tables C.8 - C.10).

Table C.8

PROMS post-hoc tests for professionals vs. non-musicians

	Pro-fessionals	**Non-musicians**					
	M (SD)	*M (SD)*	*t*	*df*[a]	*P*	*Cohens d*	
PROMS							
Pitch	**0.27 (0.06)**	**0.18 (0.06)**	6.25	75.77	<.001	1.43 [0.93, 1.94]	***
Melody	**0.23 (0.08)**	**0.07 (0.08)**	9.42	75.95	<.001	2.16 [1.59, 2.72]	***
Timbre	**0.32 (0.08)**	**0.26 (0.09)**	2.99	73.64	.004	0.70 [0.22, 1.16]	**
Rhythm	**0.33 (0.08)**	**0.27 (0.08)**	3.52	75.96	<.001	0.81 [0.34, 1.27]	***

[a] Note that original degrees of freedom were 76 but were corrected due to unequal variance.

Table C.9

PROMS post-hoc tests for professionals vs. amateurs

	Pro-fessionals	**Amateurs**					
	M (SD)	*M (SD)*	*t*	*df*[a]	*p*	*Cohens d*	
PROMS							
Pitch	**0.27 (0.06)**	**0.24 (0.07)**	2.57	87.32	.012	0.55 [0.12, 0.98]	*
Melody	**0.23 (0.08)**	**0.16 (0.10)**	4.42	95.24	<.001	0.91 [0.48, 1.33]	***
Timbre	0.32 (0.08)	0.29 (0.08)	1.72	74.69	.090	0.40 [-0.06, 0.85]	
Rhythm	0.33 (0.08)	0.32 (0.09)	0.80	84.27	.425	0.17 [-0.25, 0.60]	

[a] Note that original degrees of freedom were 126 but were corrected due to unequal variance.

Table C.10

PROMS post-hoc tests for amateurs vs. non-musicians

	Amateurs	**Non-musicians**					
	M (SD)	*M (SD)*	*t*	*df*[a]	*p*	*Cohens d*	
PROMS							
Pitch	**0.24 (0.07)**	**0.18 (0.06)**	-4.39	81.21	<.001	-0.97 [-1.43, -0.51]	***
Melody	**0.16 (0.10)**	**0.07 (0.08)**	-5.65	91.34	<.001	-1.18 [-1.62, -0.74]	***
Timbre	0.29 (0.08)	0.26 (0.09)	-1.88	62.25	.064	-0.48 [-0.98, 0.03]	
Rhythm	**0.32 (0.09)**	**0.27 (0.08)**	-3.16	80.84	.002	-0.70 [-1.15, -0.25]	***

[a] Note that original degrees of freedom were 124 but were corrected due to unequal variance.

Table C.11

Emotion classification performance – Post-hoc tests of Group x Morph Type for professionals vs. non-musicians

	Pro-fessionals	**Non-musicians**					
	M (SD)	*M (SD)*	*t*	*df*[a]	*p*	*Cohens d*	
Morph Type							
Full	**0.76 (0.06)**	**0.72 (0.07)**	-2.88	73.34	.005	-0.67 [-1.14, -0.20]	***
F0	**0.63 (0.05)**	**0.60 (0.07)**	-2.38	67.52	.020	-0.58 [-1.06, -0.09]	*
Timbre	0.41 (0.06)	0.41 (0.06)	0.21	76.00	.832	0.05 [-0.40. 0.50]	

[a] Note that original degrees of freedom were 76 but were corrected due to unequal variance.

Table C.12

Emotion classification performance – Post-hoc tests of Group x Morph Type for professionals vs. amateurs

	Pro-fessionals	**Amateurs**				
	M (SD)	*M (SD)*	*t*	*df*[a]	*p*	*Cohens d*
Morph Type						
Full	0.76 (0.06)	0.75 (0.09)	-0.99	102.5	.322	-0.20 [-0.58, 0.19]
F0	0.63 (0.05)	0.62 (0.09)	-0.78	115.7	.435	-0.15 [-0.51, 0.22]
Timbre	0.41 (0.06)	0.41 (0.07)	0.20	89.00	.840	0.04 [-0.37, 0.46]

[a] Note that original degrees of freedom were 126 but were corrected due to unequal variance.

Table C.13

Emotion classification performance – Post-hoc tests of Group x Morph Type for amateurs vs. non-musicians

	Amateurs	**Non-musicians**					
	M (SD)	*M (SD)*	*t*	*df*[a]	*p*	*Cohens d*	
Morph Type							
Full	**0.75 (0.09)**	**0.72 (0.07)**	2.04	84.38	.044	0.44 [0.01, 0.88]	*
F0	0.62 (0.09)	0.60 (0.07)	1.64	83.72	.105	0.36 [-0.08, 0.79]	
Timbre	0.41 (0.07)	0.41 (0.06)	-0.03	86.74	.975	-0.007 [-0.43, 0.41]	

[a] Note that original degrees of freedom were 124 but were corrected due to unequal variance.

Table C.14

Correlations between PROMS and VER, controlled for musical education

	$PROMS_{Avg}$	Pitch	Melody	Timbre	Rhythm
Amateurs (N = 88)					
VER_{Avg}	**.38 (.003)**	.15 (.204)	**.27 (.023)**	.22 (.066)	**.36 (.003)**
Full-Morphs	**.35 (.005)**	.14 (.212)	**.28 (.023)**	.23 (.058)	**.32 (.008)**
F0-Morphs	**.39 (.003)**	.15 (.204)	**.34 (.006)**	.24 (.053)	**.32 (.008)**
Timbre-Morphs	.18 (.124)	.08 (.503)	.05 (.673)	.08 (.499)	.22 (.062)
All groups (N = 166)					
VER_{Avg}	**.39 (<.001)**	.15 (.075)	**.30 (<.001)**	**.19 (.022)**	**.34 (<.001)**
Full-Morphs	**.35 (<.001)**	.15 (.066)	**.32 (<.001)**	**.21 (.012)**	**.25 (.003)**
F0-Morphs	**.38 (<.001)**	.11 (.174)	**.31 (<.001)**	**.19 (.021)**	**.32 (<.001)**
Timbre-Morphs	**.17 (.046)**	.08 (.310)	.03 (.666)	.07 (.382)	**.19 (.021)**

Note. VER = Vocal Emotion Recognition performance. p-values (in parenthesis) were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995).

Table C.15

Correlations between Gold-MSI and VER, controlled for musical education

	General Musicality	Active Engagement	Emotion	Singing	Perception
Amateurs (N = 88)					
VER_{Avg}	.08 (.627)	-.16 (.509)	.19 (.509)	.09 (.627)	.07 (.627)
Full-Morphs	.04 (.769)	-.14 (.509)	.13 (.533)	.07 (.627)	.02 (.829)
F0-Morphs	.16 (.509)	-.06 (.711)	.17 (.509)	.15 (.509)	.14 (.509)
Timbre-Morphs	-.09 (.628)	-.21 (.509)	.08 (.627)	-.05 (.731)	-.07 (.627)
All groups (N = 166)					
VER_{Avg}	.12 (.345)	-.08 (.525)	.05 (.718)	.15 (.261)	.12 (.350)
Full-Morphs	.12 (.345)	-.06 (.718)	.04 (.810)	.17 (.189)	.09 (.497)
F0-Morphs	.17 (.189)	-.03 (.810)	.03 (.810)	.17 (.189)	.13 (.345)
Timbre-Morphs	-.05 (.718)	-.09 (.497)	.02 (.848)	-.03 (.810)	.00 (.950)

Note. VER = Vocal Emotion Recognition performance. p-values (in parenthesis) were adjusted for multiple comparisons using the Benjamini-Hochberg correction when appropriate (Benjamini & Hochberg, 1995).

Table C.16

Correlations between PROMS and Gold-MSI

	General Musicality	Active Engagement	Formal Education	Emotion	Singing	Perception
Amateurs (N = 88)						
$PROMS_{Avg}$	**.32 (.017)**	-.02 (.892)	**.28 (.045)**	.11 (.440)	**.31 (.018)**	**.27 (.050)**
Pitch	.22 (.104)	.10 (.466)	**.26 (.050)**	.02 (.892)	.17 (.224)	.14 (.281)
Melody	**.34 (.017)**	.04 (.826)	.20 (.149)	.19 (.455)	**.39 (.006)**	**.33 (.017)**
Timbre	**.23 (.090)**	.01 (.898)	.14 (.281)	.18 (.181)	.21 (.126)	.23 (.090)
Rhythm	.16 (.256)	-.12 (.391)	.19 (.149)	.03 (.845)	.15 (.256)	.09 (.485)
All groups (N = 166)						
$PROMS_{Avg}$	**.55 (<.001)**	**.27 (<.001)**	**.52 (<.001)**	**.16 (.040)**	**.51 (<.001)**	**.50 (<.001)**
Pitch	**.44 (<.001)**	**.27 (<.001)**	**.44 (<.001)**	**.22 (.006)**	**.38 (<.001)**	**.39 (<.001)**
Melody	**.57 (<.001)**	**.34 (<.001)**	**.48 (<.001)**	**.21 (.007)**	**.54 (<.001)**	**.52 (<.001)**
Timbre	**.28 (<.001)**	.14 (.078)	**.23 (.004)**	.05 (.531)	**.29 (<.001)**	**.28 (<.001)**
Rhythm	**.24 (.003)**	.04 (.644)	**.27 (<.001)**	-.03 (.713)	**.22 (.006)**	**.21 (.009)**

Note. p-values (in parenthesis) were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995).

Acknowledgements

This thesis would not have been finished without a whole bunch of amazing people. Just like the music in all the choirs and orchestras I visited, this document was a group effort.

I would like to thank, first and foremost, Dr. Christine Nussbaum for supervising my master thesis and guiding me through the whole process with admirable patience and endless support. Your enthusiasm and expertise in the field of music and voice research is inspiring and I am thankful for the opportunity to learn and study under you. I also thank Professor Schweinberger, who generously provided knowledge and expertise in and outside of seminars and gave me the opportunity to conduct this research.

This thesis would definitely not have happened without the many musicians participating in the study. Special thanks to all the singers and instrumentalists for their time, their interest and their clever questions.

More generally, I thank with all my heart my parents and sister for supporting me through yet another degree. Without your help and love I wouldn't have done it. The same applies to my friends: you offered to join choirs just to be able to participate, patiently listened to my rambling and doubts, gave advice with humour and kindness, and I will forever be grateful.

Finally, I thank Alex, for only sometimes making me leave the home office when he had a work call, for reminding me that I'm not stupid time and time again and for unwaveringly believing in me even when I did not.

Ehrenwörtliche Erklärung

1. Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.
   Ich trage die Verantwortung für die Qualität des Textes sowie die Auswahl aller Inhalte und habe sichergestellt, dass Informationen und Argumente mit geeigneten wissenschaftlichen Quellen belegt bzw. gestützt werden. Die aus fremden oder auch eigenen, älteren Quellen wörtlich oder sinngemäß übernommenen Textstellen, Gedankengänge, Konzepte, Grafiken etc. in meinen Ausführungen habe ich als solche eindeutig gekennzeichnet und mit vollständigen Verweisen auf die jeweilige Quelle versehen. Spezifische Teile der Methoden (2.2 *Stimuli*; 2.3 *Design*, Beschreibungen des Experiments und der Fragebögen; 2.4 *Data Analysis*, Absätze eins und zwei) wurden wörtlich aus der Studie von Nussbaum, Schirmer und Schweinberger (2023b) übernommen. Dies ist im Text an den entsprechenden Stellen vermerkt und geschah mit Zustimmung der Autor:innen. Alle weiteren Inhalte dieser Arbeit ohne entsprechende Verweise stammen im urheberrechtlichen Sinn von mir.
2. Ich weiß, dass meine Eigenständigkeitserklärung sich auch auf nicht zitierfähige, generierende KI- Anwendungen (nachfolgend „generierende KI“) bezieht.
   Mir ist bewusst, dass die Verwendung von generierender KI unzulässig ist, sofern nicht deren Nutzung von der prüfenden Person ausdrücklich freigegeben wurde (Freigabeerklärung). Sofern eine Zulassung als Hilfsmittel erfolgt ist, versichere ich, dass ich mich generierender KI lediglich als Hilfsmittel bedient habe und in der vorliegenden Arbeit mein gestalterischer Einfluss deutlich überwiegt. Ich verantworte die Übernahme der von mir verwendeter maschinell generierter Passagen in meiner Arbeit vollumfänglich selbst.
   Für den Fall der Freigabe der Verwendung von generierender KI für die Erstellung der vorliegenden Arbeit wird eine Verwendung in einem gesonderten Anhang meiner Arbeit kenntlich gemacht.
   Dieser Anhang enthält eine Angabe oder eine detaillierte Dokumentation über die Verwendung generierender KI gemäß den Vorgaben in der Freigabeerklärung der prüfenden Person.
   Die Details zum Gebrauch generierender KI bei der Erstellung der vorliegenden Arbeit inklusive Art, Ziel und Umfang der Verwendung sowie die Art der Nachweispflicht habe ich der Freigabeerklärung der prüfenden Person entnommen.
3. Ich versichere des Weiteren, dass die vorliegende Arbeit bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt wurde oder in deutscher oder einer anderen Sprache als Veröffentlichung erschienen ist.
4. Mir ist bekannt, dass ein Verstoß gegen die vorbenannten Punkte prüfungsrechtliche Konsequenzen haben und insbesondere dazu führen kann, dass meine Prüfungsleistung als Täuschung und damit als mit „nicht bestanden“ bewertet werden kann. Bei mehrfachem oder schwerwiegendem Täuschungsversuch kann ich befristet oder sogar dauerhaft von der Erbringung weiterer Prüfungsleistungen in meinem Studiengang ausgeschlossen werden.

Ort und Datum | Unterschrift